LE SCIENZE
edizione italiana di SCIENTIFIC AMERICAN

numero 184
dicembre 1983
anno XVI
volume XXXI

Negoziati bilaterali e corsa agli armamenti

Una rassegna storica delle posizioni degli USA e dell'URSS, nei più importanti colloqui per il controllo degli armamenti, fa pensare che ciascuna parte abbia problemi particolari nel trattare con l'altra

di Herbert F. York

Perché è tanto difficile per gli Stati Uniti e l'Unione Sovietica negoziare accordi reciprocamente vantaggiosi sul controllo degli armamenti? L'interrogativo si pone nella maniera più esplicita nel contesto dei prolungati colloqui bilaterali tenuti a Ginevra sulla limitazione delle armi nucleari a raggio intermedio in Europa e sulla riduzione degli armamenti nucleari strategici, o a lungo raggio, in generale. Si pone però nella stessa misura anche per altri tentativi - passati, presenti e futuri - delle due superpotenze di fare ricorso a mezzi diplomatici diretti per rallentare la corsa agli armamenti e contribuire a ridurre le probabilità di una guerra nucleare. Nelle posizioni negoziali delle due parti esistono, a mio avviso, alcune importanti idiosincrasie che rendono particolarmente difficile per entrambe concludere tali accordi. Per illustrare quello che intendo dire passerò brevemente in rassegna i precedenti storici e lo stato attuale di parecchie linee d'azione di primaria importanza dei negoziati per il controllo degli armamenti. A questo proposito attingerò in parte alla mia esperienza personale di partecipante a tali sessioni, e in particolare a quelle dell'ultima serie delle più lunghe di tutte le discussioni sul controllo degli armamenti nucleari: gli sporadici tentativi venticinquennali di arrivare a una messa al bando totale degli esperimenti nucleari.

Il primo passo concreto verso il controllo della corsa agli armamenti nucleari fu la moratoria degli esperimenti nucleari alla quale tanto gli Stati Uniti quanto l'Unione Sovietica si attennero dal 1958 al 1961. Due fattori principali di grande rilievo erano alla base di questo risultato. Uno era di carattere esterno: la preoccupazione pubblica, espressa in termini sempre più pressanti sia negli Stati Uniti sia nell'Unione Sovietica, per i pericoli della ricaduta radioattiva degli esperimenti nucleari in superficie, preoccupazione suscitata in gran parte dall'esposizione fortuita alla radioattività causata dall'esperimento Bravo compiuto dagli americani nel Pacifico nel 1954. L'altro fattore era di ordine interno: la crescente preoccupazione del presidente Eisenhower e di alcuni suoi consiglieri (e a quanto pare anche del premier Kruscev e di alcuni suoi consiglieri) in merito a dove stava portando la corsa agli armamenti nucleari, preoccupazione a cui si univa la determinazione da parte di entrambi i paesi di trovare e compiere un primo passo appropriato in questa direzione. La moratoria non fu il risultato di un negoziato bilaterale, ma si reggeva soltanto ed esclusivamente su un paio di dichiarazioni pubbliche unilaterali più o meno dello stesso tenore, del tipo «ci tratterremo da ulteriori esperimenti nucleari se voi farete altrettanto».

Secondo il presidente Eisenhower lo scopo principale di una moratoria era quello di creare un clima politico favorevole all'avvio di un negoziato di un trattato particolareggiato sul tema degli esperimenti. Purtroppo però, quando erano ancora a una fase iniziale, i negoziati arrivarono a un punto morto su una questione che da allora ha continuato a ostacolare ogni tentativo di negoziare accordi bilaterali di questo genere sul controllo degli armamenti. In breve, i negoziatori americani non erano soddisfatti delle procedure di controllo che i negoziatori sovietici erano disposti a discutere, e i sovietici a loro volta accusavano gli americani di essere interessati in realtà soltanto a spiare e di pretendere troppo in fatto di intrusioni nella loro sovranità territoriale. In questo caso la questione faceva perno sul problema di rilevare e identificare le esplosioni sperimentali sotterranee. Gli Stati Uniti sostenevano che il problema poteva essere affrontato adeguatamente solo sulla base di un sistema di ispezioni *in loco* abbastanza intrusive e obbligatorie, mentre l'Unione Sovietica non era disposta ad accettare intese di questo genere.

Quando i negoziati erano in corso già da più di un anno, il presidente Eisenhower si mostrò insoddisfatto del loro andamento e, in risposta in parte alla pressione delle critiche che gli venivano rivolte all'interno della sua stessa amministrazione, annunciò che gli Stati Uniti non erano più vincolati dal loro impegno di non compiere esperimenti, ma che non avrebbero ripreso gli esperimenti stessi senza preavviso. Parecchi giorni dopo, il premier Kruscev dichiarò che in tal caso anche l'Unione Sovietica non era più vincolata dal proprio impegno a non compiere esperimenti, ma che non avrebbe rico-

minciato a farne, a meno che le nazioni occidentali non si fossero avviate per prime su questa strada. Alcuni mesi dopo, agli inizi del 1960, la Francia effettuò il suo primo esperimento nucleare. Ciononostante, né gli Stati Uniti né l'Unione Sovietica ne seguirono immediatamente l'esempio. Alla fine, più di un anno dopo, l'URSS diede il via di punto in bianco a una serie di esperimenti nucleari di rilievo, ponendo fine in tal modo alla moratoria bilaterale.

Questo comportamento dell'Unione Sovietica è stato citato spesso come esempio della «perfidia comunista». Si trattò certo, a mio parere, di una mossa sbagliata, ma non perfida, perché a quel tempo non esisteva alcun accordo - tacito o esplicito - di astenersi da tali esperimenti. In ogni caso gli Stati Uniti non tardarono molto a rispondere con una serie di propri esperimenti. Alla fine, come risultato dell'allarme suscitato dalla crisi dei missili cubani dell'ottobre del 1962, i negoziati furono ripresi, e nel 1963 il presidente Kennedy e il premier Kruscev elusero il problema del controllo di una messa al bando degli esperimepti sotterranei negoziando e firmando il Limited Test Ban Treaty, che proibiva gli esperimenti nell'atmosfera, nello spazio esterno e sott'acqua, ma permetteva di continuare quelli nel sottosuolo.

Sulla scia di questo sia pur limitato successo gli Stati Uniti e l'Unione Sovietica continuarono a esplorare la possibilità di una messa al bando di tutti gli esperimenti nucleari, senza però riuscire a risolvere la questione fondamentale del controllo. Alla fine comunque furono prese delle misure parziali aggiuntive, fra cui il Threshold Test Ban Treaty del 1974 e il Peaceful Nuclear Explosions Treaty del 1976. In più furono firmati, entrambi nel 1968, due trattati molto importanti e strettamente collegati a questo problema: il Non-Proliferation Treaty e quello (il Trattato di Tlatelolco) che istituiva una zona denuclearizzata in America Latina. La meta di una bando totale agli esperimenti è rimasta peraltro irraggiungibile.

Nel 1977 il presidente Carter incluse il negoziato per una messa al bando totale degli esperimenti fra le sue massime priorità in fatto di controllo degli armamenti. A quel tempo la posizione sovietica riguardo alle ispezioni *in loco* si era evoluta, al punto che l'Unione Sovietica era disposta ad accettare una forma volontaria di ispezione di questo genere; intanto anche la posizione americana si era evoluta, al punto che gli Stati Uniti erano disposti ad accettare una forma accuratamente circoscritta di ispezioni *in loco* volontarie anziché obbligatorie. I sovietici inoltre manifestarono la propria disponibilità ad accettare sul loro territorio un numero notevole di «stazioni sismiche nazionali» appositamente progettate e costruite. Queste stazioni, in numero da 10 a 15, avrebbero dovuto essere costruite secondo caratteristiche concordate e fornite di sistemi crittologici tali da garantire la continuità e l'invariabilità del flusso dei dati relativi.

I particolari di un trattato erano stati elaborati solo a metà quando avvenimenti esterni rallentarono i negoziati in misura tale da far diventare impossibile completarne il cammino prima della fine del mandato del presidente Carter. Tra i fattori esterni vi furono impreviste difficoltà nella seconda fase dei colloqui allora in corso sulla limitazione delle armi strategiche (SALT 2), il sequestro di ostaggi americani in Iran e l'intervento sovietico in Afghanistan. Una settimana dopo le elezione del 1980, la serie Carter di negoziati sul bando agli esperimenti fu aggiornata a tempo indeterminato.

Appena entrata in carica, l'amministrazione Reagan decise subito di non riprendere i negoziati, ma continuò a discutere per quasi un anno e mezzo le ragioni per non farlo. Un gruppo sosteneva che l'amministrazione avrebbe dovuto semplicemente dichiarare che, finché gli Stati Uniti si basavano sulle armi nucleari quale elemento importante della propria strategia difensiva, sarebbe stato necessario continuare a sperimentarle e, pertanto, una messa al bando totale degli esperimenti non sarebbe stata nell'interesse del paese per il futuro prevedibile. Un secondo gruppo sosteneva che il problema principale era costituito dal fatto che non esisteva ancora nessun sistema idoneo per controllare un bando agli esperimenti sotterranei (con il sottinteso che, se fossero esistiti, gli Stati Uniti sarebbero stati disposti forse a negoziare un accordo). Alla fine il dibattito interno vide prevalere il secondo gruppo, e l'attuale posizione ufficiale è che una messa al bando totale degli esperimenti rimane un «obiettivo a lunga scadenza» degli Stati Uniti, ma che «attualmente le condizioni internazionali non sono propizie per un'azione immediata su questo meritevole progetto».

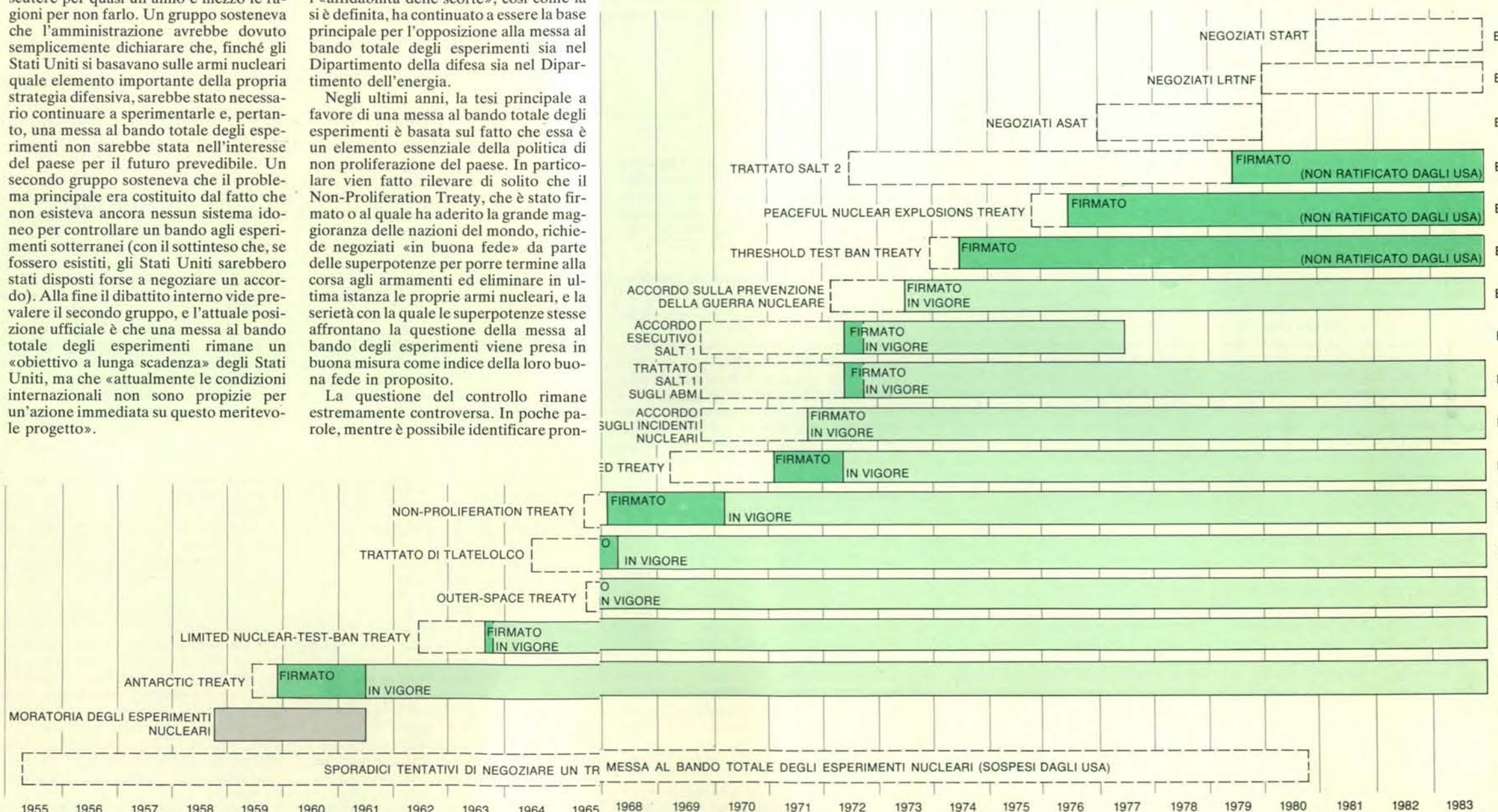
In anni recenti, compreso il periodo della presidenza Carter, i capi di stato maggiore congiunti hanno sostenuto con vigore e coerenza che una messa al bando totale degli esperimenti non sarebbe stata nei migliori interessi degli Stati Uniti, fosse o non fosse possibile sottoporla a controllo (e secondo loro non era possibile). In particolare, affermavano che, finché la nazione terrà una scorta di armi nucleari, sarà necessario condurre esperimenti almeno occasionali per essere sicuri che le armi delle scorte siano sempre in condizioni di efficienza. I capi di stato maggiore congiunti basavano questa posizione sul parere dei loro consiglieri della Defense Nuclear Agency e sui consigli della maggior parte degli esperti dei laboratori di armi nucleari. La questione dell'«affidabilità delle scorte», così come la si è definita, ha continuato a essere la base principale per l'opposizione alla messa al bando totale degli esperimenti sia nel Dipartimento della difesa sia nel Dipartimento dell'energia.

Negli ultimi anni, la tesi principale a favore di una messa al bando totale degli esperimenti è basata sul fatto che essa è un elemento essenziale della politica di non proliferazione del paese. In particolare vien fatto rilevare di solito che il Non-Proliferation Treaty, che è stato firmato o al quale ha aderito la grande maggioranza delle nazioni del mondo, richiede negoziati «in buona fede» da parte delle superpotenze per porre termine alla corsa agli armamenti ed eliminare in ultima istanza le proprie armi nucleari, e la serietà con la quale le superpotenze stesse affrontano la questione della messa al bando degli esperimenti viene presa in buona misura come indice della loro buona fede in proposito.

La questione del controllo rimane estremamente controversa. In poche parole, mentre è possibile identificare prontamente, per mezzo di sistemi di telerilevamento, eventuali esplosioni nucleari di carica elevata o modesta, quelle di cariche piccolissime sfuggono a qualsiasi possibilità di rilevamento. La linea di demarcazione che separa le due categorie è peraltro indistinta, e ciò porta a interpretazioni molto diverse dei dati, a seconda della predisposizione della persona che li interpreta.

Sin dall'inizio dell'era nucleare il punto di vista prevalente sia negli Stati Uniti sia nell'Unione Sovietica è che gli interessi di tutti e due i paesi, e anzi del mondo intero, sarebbero serviti meglio se non ci fossero altre potenze nucleari, o quanto meno se ce ne fosse il minor numero possibile. Questo è uno dei rari casi in cui tanto gli americani quanto i sovietici sembrano rendersi conto perfettamente che i loro interessi sono esattamente paralleli.

Gli Stati Uniti, a volte unilateralmente e a volte di concerto con altre nazioni, hanno concepito numerose politiche intese a favorire l'obiettivo della non proliferazione delle armi nucleari. Fra le misure adottate sono da annoverare la creazione dell'International Atomic Energy Agency, l'approvazione da parte del Congresso del Nonproliferation Act del 1978 e numerose iniziative diplomatiche tendenti fra l'altro a limitare rigorosamente il ritrattamento del combustibile esaurito dei reattori nucleari e a porre restrizioni alla vendita di attrezzature per la produzione di energia nucleare. Non meno importante, è il Non-Proliferation Treaty del 1968, entrato in vigore nel 1970.

In sostanza, il Non-Proliferation Treaty può essere considerato un tentativo di dividere le nazioni del mondo in due categorie: una, costituita da quei paesi che già disponevano di armi nucleari e che si erano impegnati a non aiutare nessun altro a procurarsele, e l'altra, composta da quelle nazioni che a quel tempo non avevano armi nucleari e promettevano di rinunciarvi per sempre. In cambio della promessa di non acquistare armi nucleari, i paesi non nucleari pretesero dalle nazioni firmatarie che

Nell'illustrazione è riportata la cronologia dei più importanti negoziati per il controllo degli armamenti tenutisi dal 1955. La parte tratteggiata di ogni istogramma indica la durata delle trattative, la parte in colore chiaro corrisponde al tempo intercorso tra la firma del trattato e la sua ratifica, mentre la parte in colore intenso indica il periodo durante il quale l'accordo è in vigore. Le lettere maiuscole all'estrema destra indicano se l'accordo è multilaterale (*M*) o bilaterale (*B*) tra Stati Uniti e Unione Sovietica. Tre trattati bilaterali non sono stati ancora ratificati dal Senato americano: il Threshold Test Ban Treaty (firmato dal presidente Nixon nel 1974), il Peaceful Nuclear Explosions Treaty (firmato dal presidente Ford nel 1976) e il SALT 2 Treaty (firmato dal presidente Carter nel 1979). USA e URSS sono impegnati attualmente in due negoziati per il controllo degli armamenti nucleari: i colloqui per la riduzione delle armi strategiche (START) e la discussione sulla limitazione delle armi nucleari a raggio intermedio in Europa. (Queste armi sono note tecnicamente come forze nucleari «da teatro» a lungo raggio, o LRTNF.) I negoziati intesi a raggiungere una messa al bando totale degli esperimenti nucleari e un accordo per prevenire l'ulteriore sviluppo e spiegamento di armi antisatellite sono stati a quanto pare abbandonati dall'amministrazione Reagan.

già ne disponevano due assicurazioni. La prima era che «chi le aveva» avrebbe aiutato «chi non le aveva» ad acquisire la tecnologia necessaria per permettere loro di trarre pieno profitto dalle applicazioni pacifiche dell'energia nucleare; l'altra era che «chi le aveva» avrebbe avviato serie trattative per porre fine alla corsa agli armamenti nucleari ed eliminare da ultimo le proprie stesse armi nucleari. Il trattato prevede anche una revisione della situazione globale ogni cinque anni. Nella prima e nella seconda conferenza di revisione (nel 1975 e nel 1980) molti paesi partecipanti lamentarono il fatto che nessuna delle due superpotenze si atteneva agli obblighi particolari sopra indicati. Nonostante queste lamentale, non si è avuta nessuna defezione dal trattato e, cosa ancora più importante, dalla firma del trattato in poi non c'è stata nessuna ulteriore proliferazione di armi nucleari.

Vi sono ovviamente casi in cui un'ulteriore proliferazione potrebbe avvenire subito (India, Pakistan, Israele, Repubblica Sudafricana, Argentina e Brasile, per esempio), ma fino a questo momento anche quei paesi che a quanto pare sono andati molto avanti sulla strada della capacità nucleare si sono trattenuti dal compiere esperimenti o dall'istituirsi apertamente in qualsiasi altro modo quali stati detentori di armi nucleari. (Questa osservazione vale anche per l'India, che 10 anni fa fece esplodere un ordigno nucleare, ma che non ha creato ancora una forza nucleare.) Certo le politiche e le azioni cui si è fatto cenno sin qui, compreso in particolare il Non-Proliferation Treaty, spiegano in maniera preponderante il lungo e quasi inatteso ritardo della comparsa di altre potenze nucleari oltre alle cinque che dispongono di armi nucleari sin da quando la Cina divenne membro del «club» nel 1964.

Nonostante questo risultato di successo generale, va notato che alcune nazioni molto importanti, fra cui Cina, Francia, Cuba, India, Argentina, Brasile e Israele, hanno rifiutato di firmare il Non-Proliferation Treaty. Nel caso della Francia e della Cina l'omissione è particolarmente grave, ma per lo meno ognuno di questi due paesi, pur rifiutando, per ragioni di natura politica, di aderirvi formalmente, si è impegnato a non venir meno allo spirito del trattato.

Agli inizi del suo mandato presidenziale, Johnson propose un congelamento dell'ulteriore sviluppo o spiegamento di vettori strategici. Dopo una breve discussione della proposta ai colloqui di Ginevra sugli armamenti, il presidente Johnson s'incontrò con il premier Kosygin a Glassboro, nel New Jersey, dove la questione fu discussa in sede privata. Fu in quest'occasione che il Segretario alla difesa, Robert S. McNamara, anch'egli presente, avanzò formalmente per la prima volta l'idea di iniziare tale congelamento proibendo lo spiegamento di sistemi di missili antibalistici (ABM). Il premier Kosygin respinse l'idea sostenendo che le armi ABM erano puramente difensive, mentre, tutto sommato, erano soltanto le armi offensive quelle che minacciavano la vita di molti milioni di persone. Il Segretario alla difesa McNamara e altri americani continuarono peraltro a sostenere in varie riunioni pubbliche e private che lo sviluppo e lo spiegamento di sistemi ABM stimolavano la corsa agli armamenti esattamente come lo sviluppo e lo spiegamento di sistemi offensivi, e ciò a causa di un «ciclo di azione e reazione» in cui lo sviluppo di nuove armi difensive promuoveva lo sviluppo di nuove armi offensive. Alla fine i sovietici accettarono questo punto di vista.

I tentativi del presidente Johnson di avviare formali negoziati bilaterali sulla limitazione delle armi strategiche furono frustrati dall'intervento sovietico in Cecoslovacchia nel 1968. Soltanto dopo che il presidente Nixon entrò in carica nel 1969 e le acque internazionali si placarono fu possibile dare il via a quelli che oggi sono noti come colloqui SALT. La prima serie di questi colloqui (SALT 1), conclusa nel 1972, ebbe due risultati: un trattato che limitava rigorosamente lo spiegamento di sistemi ABM e un accordo esecutivo che bloccava temporaneamente lo spiegamento di missili offensivi con base a terra e in mare, congelandolo al numero di quelli già spiegati più quelli il cui spiegamento era in corso al momento della firma del trattato. Il risultato finale fu un equilibrio strategico approssimativo che era inteso non come fine a se stesso, ma come base per ulteriori negoziati, il cui fine ultimo era quello di creare una situazione di effettiva parità generale a un livello molto più basso di spiegamento totale.

Uno dei principali problemi residui degli accordi SALT 1 fu il fatto che l'Unione Sovietica rimaneva con un numero considerevole di missili molto grandi (designati SS-18 dalle fonti occidentali) mentre gli Stati Uniti non disponevano di missili di dimensioni paragonabili. Agli inizi degli anni settanta, prima del completo sviluppo e dello spiegamento su vasta scala dei MIRV (Multiple Independently Targetable Reentry Vehicles, ossia missili a testate multiple a bersaglio indipendente) estremamente precisi, questo squilibrio non appariva grave alla maggioranza degli osservatori, ma in anni recenti, con il perfezionamento di sistemi MIRV a elevato numero di testate e con il conseguimento di un altissimo grado di precisione da parte non solo degli Stati Uniti, ma anche dell'Unione Sovietica, la questione degli SS-18 è considerata ormai un po' dappertutto una faccenda seria. I nuovi sviluppi aumentano la vulnerabilità della componente con base a terra della «triade» americana, la parte cioè su cui la politica difensiva degli Stati Uniti fa maggiore affidamento.

Il passo successivo dopo i SALT 1 fu l'accordo di Vladivostok del 1974, firmato dal presidente Ford e dal presidente Breznev. In sostanza tale accordo era inteso a stabilire direttive generali per i negoziati particolareggiati SALT 2, che avevano già avuto inizio a Ginevra. Le direttive in questione prevedevano un «ugual limite aggregato» di 2400 per i vettori e un «ugual limite aggregato» di 1320 per i sistemi MIRV. Lo sforzo di trasformare prontamente le direttive di Vladivostok in un trattato formale si scontrò con problemi di natura sia tecnica sia politica. I problemi tecnici riguardavano i missili da crociera e il bombardiere sovietico «Backfire»; i problemi politici erano dovuti alla crescente disillusione a proposito della distensione e alla sfida alla designazione del presidente Ford lanciata dalla destra del Partito repubblicano.

Nel 1977 i negoziati SALT 2 diventarono l'elemento principale della politica del presidente Carter in fatto di controllo degli armamenti. Sulle prime Carter propose alcune riduzioni del numero dei sistemi spiegati, riduzioni che superavano gli aggregati di Vladivostok in quanto prevedevano un taglio del 50 per cento per i missili pesanti sovietici (le cui versioni moderne erano completamente proibite agli Stati Uniti), nonché limiti agli esperimenti, ai miglioramenti e al numero dei missili balistici intercontinentali (ICBM) dotati di MIRV. Queste ultime disposizioni erano intese a prevenire un ulteriore aggravamento del problema della vulnerabilità degli ICBM e a ridurre in genere le pressioni di natura tecnologica sulla corsa agli armamenti. L'URSS insistette però per rimanere entro la cornice di Vladivostok. I negoziati SALT 2 furono portati a compimento a Ginevra e in una serie di incontri ad alto livello nel corso dei due anni successivi, e il trattato che ne conseguì fu firmato a Vienna nel 1979 dal presidente Carter e dal presidente Breznev.

Il Trattato SALT 2, che conservava sostanzialmente la cornice di Vladivostok, poneva a ognuna delle due parti un limite complessivo di 2250 per i vettori nucleari strategici di ogni genere, ed entro tale limite prevedeva una serie di sottolimiti, i quali fissavano rispettivamente un tetto di 1320, 1200 e 820 unità ai sistemi di qualsiasi genere dotati di MIRV (compresi i bombardieri pesanti con missili da crociera a lungo raggio), ai missili balistici dotati di MIRV e agli ICBM dotati di MIRV. Il trattato limitava inoltre ognuna delle due parti a un solo tipo nuovo di ICBM, vietava cambiamenti di rilievo nei sistemi esistenti e poneva limiti specifici al numero dei veicoli di rientro che potessero essere installati sull'unico ICBM nuovo e su ogni tipo di ICBM esistente. Limiti temporanei ai missili da crociera lanciati da terra (GLCM), ai missili da crociera lanciati dal mare (SLCM) e agli ICBM mobili erano intesi inoltre a dare il tempo necessario per negoziare accordi a lunga scadenza su tali sistemi.

Molte delle critiche avanzate negli Stati Uniti al Trattato SALT 2 scaturirono dal diffuso senso di frustrazione di fronte alle difficoltà internazionali del paese, dagli esagerati timori di inferiorità nucleare (stimolati in patria dalle critiche della destra) e dalla mancanza di fiducia nella competenza del presidente Carter relativamente alla sicurezza in generale e al suo impegno in programmi militari in particolare. Inoltre l'imminenza delle elezioni presidenziali rendeva riluttanti i senatori repubblicani a offrire una grossa vittoria in politica estera a un avversario indebolito. Nonostante la generale soddisfazione per i termini complessivi dell'accordo, c'erano nel trattato parecchi punti particolari che sollevavano gravi interrogativi nella mente di quei senatori e di altre personalità politiche che erano estremamente caute a proposito dell'intera faccenda del controllo degli armamenti.

Uno di questi interrogativi riguardava il fatto che, al pari dei SALT 1, i SALT 2 non facevano nulla per attenuare il problema dei missili pesanti oltre a porre un limite piuttosto alto al numero di testate che ogni singolo tipo di missile poteva trasportare. Una seconda questione riguardava il bombardiere Backfire, un aereo che in certe circostanze, sia pure inconsuete, è in grado di raggiungere gli Stati Uniti. Alcuni osservatori americani affermarono pertanto che il Backfire era un bombardiere intercontinentale e, come tale, avrebbe dovuto rientrare nei limiti fissati. L'URSS sosteneva che la missione di volo del Backfire era solo quella di un bombardiere a raggio intermedio, e quindi l'aereo non avrebbe dovuto rientrare nei limiti suddetti, tenuto conto soprattutto del fatto che a Vladivostok erano stati esclusi, nonostante le forti obiezioni dell'Unione Sovietica, gli aerei americani a medio raggio dislocati in Europa in grado di raggiungere l'URSS. Una terza questione riguardava la codificazione di certi dati degli esperimenti trasmessi dai missili sovietici durante i voli sperimentali. Quest'ultimo è un argomento piuttosto arcano che non può essere affrontato utilmente in una discussione non soggetta alle norme relative alla sicurezza. Basti dire che la maggior parte dei professionisti coinvolti a quel tempo nelle trattative erano convinti che la faccenda fosse stata trattata in maniera adeguata.

Da parte sovietica la questione principale sollevata dopo l'incontro di Vladivostok riguardava i missili da crociera, che l'Unione Sovietica voleva limitare nettamente. (Per evitare limiti ai missili da crociera sovietici esistenti, i negoziatori sovietici insistettero sul fatto che ogni limite doveva valere soltanto per i missili con un raggio d'azione di oltre 600 chilometri.) La questione assunse un carattere ancora più pressante con la discussione fra gli stati membri della North Atlantic Treaty Organization (NATO) sul possibile spiegamento in Europa di versioni di tali sistemi con base di lancio a terra e con la decisione del presidente Carter di eliminare il bombardiere B-1 e di accentuare lo spiegamento di missili da crociera lanciati dall'aria (ALCM) installati su bombardieri con base negli Stati Uniti continentali.

Nel caso degli ALCM a lungo raggio il problema principale era costituito non dalla messa al bando, ma dal modo in cui si sarebbero dovuti contare i missili. Un bombardiere che trasportava ALCM doveva essere contato come un solo vettore, indipendentemente dal numero dei missili che trasportava (come è di norma per i bombardieri che trasportano bombe o missili da crociera a breve raggio), oppure bisognava contare come vettore ogni singolo ALCM (come nel caso dei missili balistici lanciati da un sottomarino)? Alla fine l'URSS accettò un complesso compromesso in base al quale si sarebbero contati solo i bombardieri, a condizione che nessun tipo esistente di bombardiere potesse essere dotato di più di 20 missili da crociera e che il numero medio dei missili da crociera su tutti i bombardieri così attrezzati non superasse i 28. Se, inoltre, il numero di bombardieri dotati di missili da crociera avesse superato i 120, si sarebbe dovuto ridurre in misura corrispondente il numero dei missili dotati di MIRV.

Riguardo ai missili da crociera lanciati da terra e dal mare, gli Stati Uniti cercarono di fare in modo che non fosse fissato nessun limite, perché l'accordo non prevedeva restrizioni ad altri sistemi a raggio intermedio (come il nuovo missile sovietico SS-20); l'Unione Sovietica ne voleva invece la messa al bando totale. Alla fine ne furono permessi la sperimentazione e lo sviluppo, mentre lo spiegamento fu sospeso sino alla fine del 1981 in modo da permettere ulteriori negoziati.

Succesivamente tutti questi problemi furono risolti con soddisfazione dei due presidenti e della maggior parte dei loro consiglieri, ma continuarono a essere alla base dell'opposizione al trattato all'interno dello Stato americano e, di conseguenza, il processo di ratifica fu rinviato. Alla fine, avvenimenti esterni, come per esempio la vicenda degli ostaggi americani in Iran e l'intervento sovietico in Afghanistan, ebbero il sopravvento e misero la parola fine al tentativo di ratificare il trattato. Alcuni funzionari dell'amministrazione Carter coinvolti nel tentativo, al quale partecipavo personalmente, non hanno perso la convinzione che, nonostante le difficoltà incontrate in Campidoglio, il trattato alla fine sarebbe stato ratificato se non fosse stato per l'intervento sovietico in Afghanistan.

Durante la campagna presidenziale del 1980, Reagan e i suoi sostenitori criticarono duramente, a volte sulla base dei problemi cui ho fatto cenno, i SALT 2, definendoli «fatalmente incrinati». Subito dopo l'elezione il presidente Reagan ritirò il Trattato SALT 2 da ogni ulteriore esame da parte del Senato, ma dichiarò che gli Stati Uniti si sarebbe attenuti a esso finché vi si fosse attenuta anche l'Unione Sovietica e diede il via a una nuova serie di negoziati, chiamati START (acrostico di Strategic Arms Reduction Talks), come modo per arrivare a un trattato più gradito a lui e ai suoi sostenitori. Reagan mise in particolare rilievo anche la necessità di consistenti riduzioni del numero dei sistemi spiegati, con particolare riferimento a quelli che hanno tempi di volo molto brevi (missili balistici sia con base a terra sia con base in mare) e che quindi sono considerati particolarmente destabilizzanti.

Si è avuto notizia, in particolare, che le posizioni negoziali americane agli START prevedevano inizialmente una riduzione complessiva di oltre il 50 per cento dei vettori balistici (cosicché il numero dei missili balistici sarebbe sceso a 850, e questi 850 non avrebbero potuto portare in tutto più di 5000 veicoli di rientro), nonché una riduzione del 50 per cento anche degli ICBM pesanti. Si dice inoltre che la posizione americana preveda la necessità di rinviare a epoca più tarda l'esame dei limiti da porre ai bombardieri e ai missili da crociera. È stato detto altresì che la posizione iniziale dell'Unione Sovietica prevedeva la riduzione a 1800, per ognuna delle due parti, del numero dei vettori di ogni tipo (compresi i bombardieri), una cifra proposta per la prima volta dal presidente Carter nel 1977 e respinta energicamente a quel tempo dall'URSS. Poco tempo fa il presidente Reagan ha detto che avrebbe mostrato maggiore flessibilità su alcuni di questi problemi e ha dichiarato specificamente che il numero totale dei missili balistici potrebbe essere fissato a mezza strada tra la cifra iniziale sovietica e quella americana.

Dal punto di vista dell'Unione Sovietica il difetto principale della posizione americana è che essa mette in evidenza unicamente quei settori in cui sono in vantaggio i sovietici, mentre ignora completamente quelli in cui sono gli Stati Uniti a essere in vantaggio. Dal punto di vista americano il difetto principale della posizione sovietica è che essa continua evidentemente a ignorare l'importantissimo problema dei missili pesanti.

Oltre ai problemi cui fin qui ho fatto cenno e che da qualche anno sono alla base di controversie sia all'interno della comunità americana addetta alla sicurezza nazionale sia tra gli Stati Uniti e l'Unione Sovietica, altre due questioni hanno assunto sempre maggiore importanza. Una è la questione dei missili mobili, sorta soprattutto a causa di alcuni dei modi che sono stati proposti per lo spiegamento dell'MX, il progettato successore dell'ICBM Minuteman. L'altra è la questione della «capacità di ricaricamento». È facile contare i silos missilistici, ma non è altrettanto facile contare i missili stessi. Se quindi l'una o l'altra delle due parti dovesse costruire e tenere di scorta un numero di missili paragonabile al numero dei silos, si avrebbe come conseguenza una nuova, grave fonte di potenziale instabilità. Resta da vedere se la situazione attuale arriverà a un punto morto o se ci saranno progressi sulla via della soluzione di questi problemi, ma per quel che mi riguarda non sono ottimista.

Si è parlato molto negli Stati Uniti a livello pubblico delle concessioni americane - vere e immaginate - fatte nel corso dei negoziati post-bellici con l'Unione Sovietica per il controllo degli armamenti. La storia popolare dell'Occidente non dice praticamente nulla però sulle concessioni sovietiche, sebbene l'URSS ne abbia fatte molte di importanti. Una di queste concessioni riguarda la definizione dell'espressione «sistema di vettori strategici».

Gli americani hanno sempre sostenuto che un sistema del genere è costituito da vettori strategici che, spiegati sul territorio nazionale o in mare, possono raggiungere, dal sito del loro spiegamento, il territorio nazionale della parte avversa. Per i sovietici, invece, si tratta di un sistema che, dal sito in cui è spiegato, in qualunque punto si trovi il sito stesso, è in grado di raggiungere il territorio nazionale dell'una o dell'altra parte.

Questa differenza di punti di vista è dovuta alla situazione geopolitica molto diversa delle due superpotenze e si presenta più nettamente nel caso dei «sistemi con base avanzata», vale a dire di quei sistemi americani a medio raggio che attualmente sono spiegati in Europa e che solo così possono raggiungere l'Unione Sovietica. Gli Stati Uniti si sono sempre rifiutati di contare questi sistemi nei totali SALT, mentre l'Unione Sovietica ha sempre sostenuto con fermezza che invece bisognava includerli. Fino a questo momento i sovietici hanno accettato il punto di vista americano in proposito. (A quanto pare questi sistemi sono inclusi attualmente nei colloqui organizzati separatamente a Ginevra sulle armi a raggio intermedio.)

Una seconda concessione sovietica riguarda le forze nucleari inglesi e francesi. L'Unione Sovietica sostiene fermamente che i vettori di questi due paesi della NATO dovrebbero essere inclusi nei totali SALT. Gli Stati Uniti sostengono invece con altrettanta fermezza che i SALT sono negoziati rigorosamente bilaterali e, quindi, vanno prese in considerazione soltanto le forze sovietiche e americane. Nell'accordo SALT 1 i sovietici dicevano che avrebbero accettato la posizione americana a patto che, insieme, francesi e inglesi non mettessero in mare più di nove sommergibili lanciamissili balistici (più precisamente, che il numero totale dei sommergibili messi in mare da Stati Uniti, Inghilterra e Francia non superasse i 50). Gli USA rifiutarono di accettare anche questo limite, ma la faccenda rimane controversa, in quanto quel numero non è stato superato.

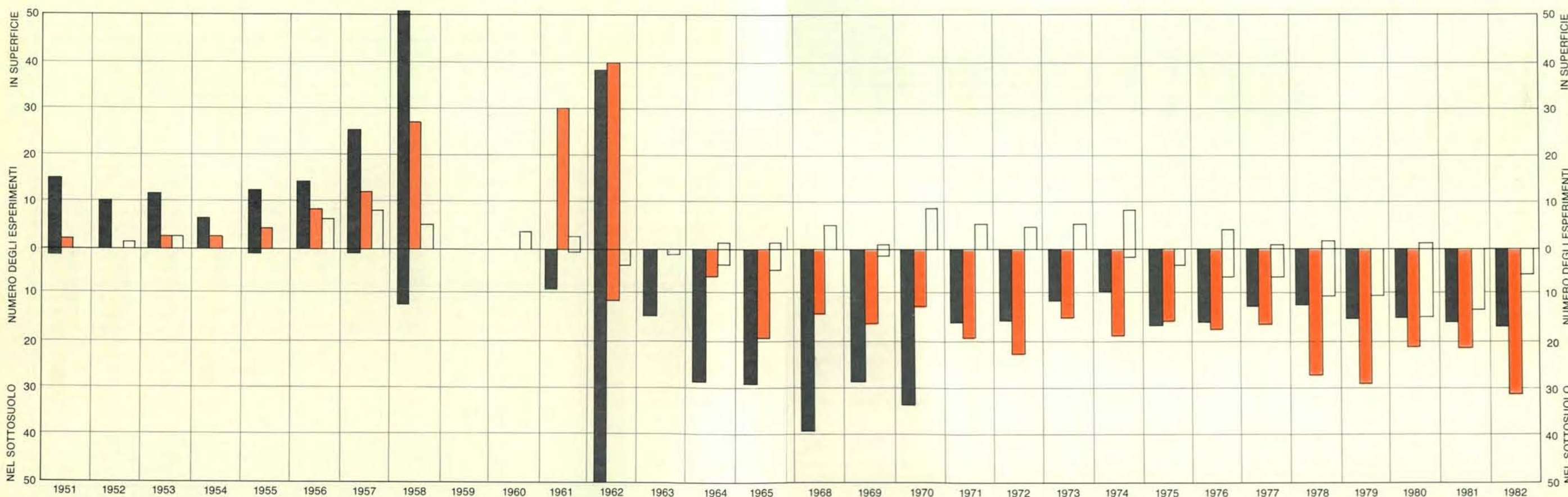
Una terza concessione di rilievo da parte dell'Unione Sovietica riguarda la procedura per contare gli ALCM installati sui bombardieri americani. Fintantoché il numero totale di questi missili da crociera rimane inferiore a 3000 e fintantoché vengono rispettate certe regole di calcolo, sarà necessario contare e limitare nell'ambito dei «tetti» fissati dai SALT soltanto il numero dei bombardieri così attrezzati, ma non quello dei missili da crociera. Come nel caso delle concessioni americane, divulgate più diffusamente, l'Unione Sovietica a quanto pare ha fatto queste e altre concessioni in un serio spirito di compromesso al fine di produrre un risultato tale da recare giovamento a tutti.

Solo qualche settimana prima delle elezioni presidenziali del 1980, USA e URSS diedero il via a negoziati intesi a limitare o a eliminare i missili a raggio intermedio installati in Europa (noti tecnicamente in Occidente come forze nucleari «da teatro» a lungo raggio). A questi negoziati, ancora in corso, hanno contribuito fattori di ordine generale e particolare. Il fattore generale è che vi sono moltissime armi nucleari (10 000 o più) spiegate nell'Europa occidentale o puntate su di essa, eppure pochissime di queste armi sono incluse nei negoziati START. Il fattore particolare deriva dallo spiegamento sovietico dell'SS-20, la prima arma nuova nella sua categoria da una ventina d'anni a questa parte. Lo spiegamento dell'SS-20 ebbe inizio nel 1977 e, da allora, ha proceduto a ritmo costante. Nel 1978 la situazione era giunta al punto che nei paesi occidentali molte persone si convinsero della necessità di rispondere in qualche modo, se non per ragioni rigorosamente militari, almeno per ragioni politiche.

Più o meno a quell'epoca il cancelliere Schmidt della Repubblica Federale Tedesca tenne un discorso nel quale attirava in modo particolare l'attenzione su tale situazione. Schmidt affermava l'assoluta necessità di un sistema NATO con base a terra estremamente visibile che controbilanciasse più o meno l'SS-20. In risposta a questa crescente preoccupazione, il Consiglio della NATO adottò nel dicembre del 1978 quella che viene chiamata l'impostazione «a due piste». Una pista chiedeva lo sviluppo e lo spiegamento di un tale sistema NATO, mentre l'altra chiedeva negoziati intesi a rendere inutile il sistema in questione. Uno dei risultati di questa duplice impostazione fu la decisione, da parte degli Stati Uniti e degli alleati della NATO, di spiegare in Europa sulla terraferma 572 missili a raggio intermedio di due tipi nuovi: un missile da crociera lanciato da terra e il Pershing II, una versione ammodernata a raggio più lungo del missile Pershing IA. Nel momento in cui scrivo né l'uno né l'altro sono stati ancora installati, ma si suppone che il Pershing II lo sarà verso la fine di quest'anno.

L'altro risultato dell'impostazione a due piste fu l'avvio a Ginevra, verso la fine di ottobre del 1980, di negoziati rigorosamente bilaterali tra le due superpotenze. I colloqui non erano ancora quasi incominciati quando arrivarono i risultati delle elezioni americane, e tutti e due i governi si resero conto che era necessario aspettare che la polvere si posasse prima di poter approdare a qualche cosa. Dopo alcuni mesi di studio l'amministrazione Reagan riprese i negoziati e propose quella che è nota come l'«opzione zero-zero». Questa proposta prevedeva l'eliminazione di tutti i missili sovietici a medio raggio, compresi gli SS-20 (o quanto meno quelli a ovest degli Urali), in cambio della promessa da parte dei paesi occidentali di rinunciare al progettato spiegamento di una forza controbilanciante. La risposta dell'URSS, giunta agli inizi della nuova amministrazione Andropov, proponeva che gli Stati Uniti rinunciassero a spiegare missili a medio raggio e, in cambio, l'Unione Sovietica avrebbe ridotto i propri SS-20 in modo da portarli allo stesso livello numerico (162) dei missili inglesi e francesi.

Considerata in astratto, l'opzione zero-zero è chiaramente una proposta auspicabile e ragionevole. Essa elimina un'intera categoria di armi, una categoria che negli anni cinquanta avrebbe forse potuto colmare un vero divario, ma i cui compiti possono essere svolti facilmente da altre armi di cui ognuna delle due parti dispone in abbondanza. Considerata però in concreto, la proposta americana chiede all'Unione Sovietica di rinunciare ad armi già acquistate e pagate, mentre gli Stati Uniti rinunciano semplicemente a certi loro piani futuri. Perciò, per quanto possa sembrare auspicabile e ragionevole, nel pessimo clima politico attuale l'opzione zero-zero risulterà insufficiente a superare le difficoltà dovute all'asimmetria della situazione concreta. Nel momento in cui scrivo nessuna delle due parti si è staccata dalla propria posizione di partenza, anche se il presidente Reagan ha fatto sapere che da parte americana potrebbe esserci un certo spazio per il compromesso.

Dal 1958 fin verso il 1977 il metodo principale usato dagli Stati Uniti per impedire lo sviluppo e lo spiegamento di ordigni adatti per una guerra nello spazio esterno o dallo spazio esterno fu quello di usare moderazione nei propri programmi e sollecitare l'Unione Sovietica a fare altrettanto. In tutto quel periodo gli Stati Uniti spiegarono sistemi spaziali che svolgevano importanti funzioni di appoggio (quali ricognizione, sorveglianza e comunicazioni), e i militari a poco a poco si abituarono a fare sempre più affidamento su tali sistemi. Proposte di sviluppare un sistema antisatellite (ASAT) di uso generale furono avanzate frequentemente negli Stati Uniti dai militari e da organizzazioni industriali, ma furono sempre respinte dalle autorità superiori. La tesi di fondo era che le attività spaziali della nazione erano diventate tanto preziose che un mondo in cui nessuna delle due superpotenze avesse armi antisatellite era migliore di un mondo in cui tutte e due ne disponessero e che, pertanto, gli Stati Uniti non avrebbero dovuto stabilire un precedente con lo sviluppo di un proprio sistema ASAT. Ci fu solo una piccola eccezione a questa politica: lo sviluppo e l'installazione sull'isola Johnston, agli inizi degli anni sessanta, di un sistema ASAT di capacità molto limitata, perché gli Stati Uniti sospettavano che l'Unione Sovieti-

Sono riportati gli esperimenti nucleari effettuati dagli USA (*istogrammi in nero*), dall'URSS (*istogrammi in colore*) e da tutte le altre nazioni (*istogrammi in bianco*) dal 1951 a tutto il 1982. Lo spazio vuoto corrisponde alla moratoria durata dal 1958 al 1961. Il risultato del Limited Test Ban del 1963 non fu quello di diminuire il numero degli esperimenti, ma di trasferirli in gran parte nel sottosuolo. (Da allora soltanto la Francia e la Cina hanno effettuato esperimenti nell'atmosfera.) I dati sono del SIPRI (Stockholm International Peace Research Institute).

ca stesse sviluppando un sistema che comportava la messa in orbita intorno alla Terra di armi nucleari estremamente potenti. Il sistema ASAT sovietico non vide però mai la luce, e dopo solo qualche anno gli Stati Uniti disarmarono il loro sistema ASAT specializzato.

Nonostante la cautela americana, l'URSS diede il via nel 1977 a un sistema ASAT di capacità limitata di uso generale. A quanto pare però i sovietici misero questo sviluppo all'ultimo posto sulla scala delle priorità rispetto a quasi tutte le altre parti del loro programma spaziale, ed esso procedette piuttosto lentamente. Di quando in quando gli Stati Uniti facevano notare la loro moderazione in questo campo e sollecitavano l'Unione Sovietica a comportarsi nello stesso modo. Può darsi che la cosa abbia avuto qualche effetto, ma il programma sovietico continuò ugualmente.

Alla fine, nel 1977, l'amministrazione Carter decise che era necessario fare qualcosa di più specifico per rispondere allo sviluppo dell'ASAT sovietico e così fu dato il via a un programma articolato in tre direzioni. Un elemento di tale programma era la decisione di iniziare uno sforzo in piena regola per sviluppare un sistema ASAT americano. (Uno sviluppo casuale era già stato iniziato durante l'amministrazione Ford.) Un secondo elemento era un piano per esplorare e sviluppare sistemi intesi a difendere i satelliti americani da un possibile attacco di armi ASAT sovietiche. Il terzo elemento era l'avvio di negoziati intesi, se possibile, a prevenire l'ulteriore sviluppo e lo spiegamento di armi ASAT da entrambe le parti.

I primi negoziati formali sui sistemi ASAT ebbero luogo a Helsinki nel 1978. Altre due sessioni di trattative si tennero nel 1979, dopo di che i negoziati furono aggiornati a epoca indeterminata, e a tutt'oggi non sono mai stati ripresi. Il fatto di non essere approdati a nulla è dovuto a ragioni sia di carattere esterno sia di carattere interno. All'interno, il problema era duplice. Innanzitutto il Governo americano era incapace di assumere un atteggiamento unitario riguardo a quel che bisognava fare. Da un lato gli alti funzionari dell'Air Force e del Dipartimento della difesa davano così tanto valore agli elementi spaziali esistenti del paese da essere disposti a rinunciare allo sviluppo di un sistema ASAT se, così facendo, avessero avuto la sicurezza che anche i sovietici avrebbero fatto altrettanto. La posizione della Marina, d'altra parte, era che i satelliti sovietici di sorveglianza degli oceani costituivano per la flotta americana una minaccia tanto grave che il paese aveva assoluto bisogno di un proprio sistema ASAT, indipendentemente da quello che facevano i sovietici. Per di più, agli ufficiali dell'Air Force di grado medio non andava a genio l'idea di barriere «politiche» frapposte a sviluppi spaziali di qualsiasi genere. Oltre a queste differenze, il fatto che l'Unione Sovietica avesse già condotto nel decennio precedente più di una dozzina di esperimenti ASAT suscitò forti pressioni perché gli Stati Uniti svolgessero almeno un numero equivalente di esperimenti analoghi.

Il secondo problema interno dei negoziati era costituito dal fatto che anche i sovietici sembravano ambivalenti sulla questione. Nel corso delle trattative essi resero noto che a loro modo di vedere potevano esserci serie e valide ragioni per possedere una capacità ASAT e sollevarono anche una serie di interrogativi sulla navetta spaziale americana e sul ruolo che essa poteva avere in questo contesto. Oltre a tutti questi problemi che riguardavano direttamente la sostanza dei negoziati ASAT, lo stesso deterioramento generale del clima internazionale che aveva ostacolato i SALT 2 e i negoziati per la messa al bando totale degli esperimenti ebbe la sua parte nell'impedire ulteriori progressi in questo campo.

Oltre a questi fin qui futili tentativi di prevenire lo sviluppo di sistemi antisatellite, gli Stati Uniti e l'Unione Sovietica sono riusciti a negoziare in questo settore accordi più limitati. Di grande importanza sono le clausole del Limited Test Ban Treaty che proibiscono le esplosioni nello spazio esterno, l'Outer Space Treaty del 1967 e le clausole del Trattato SALT 1 che legittimano e proteggono i satelliti da ricognizione usati per controllare gli accordi raggiunti in materia di controllo degli armamenti e che vietano lo sviluppo e lo spiegamento di sistemi ABM con base nello spazio. L'Outer Space Treaty proibisce di collocare nello spazio esterno o su corpi celesti «armi di distruzione di massa». L'espressione tra virgolette viene interpretata sempre in modo da includere le armi nucleari ed escludere quelle non nucleari.

Più di recente, nel 1981, la proposta di un accordo inteso a prevenire i tentativi di una guerra nello spazio o dallo spazio fu presentata alle Nazioni Unite dal presidente Breznev. In breve, la proposta prevedeva la proibizione dello spiegamento di armi di qualsiasi genere - non solo di quelle di distruzione di massa - in orbita intorno alla Terra, su corpi celesti «o nello spazio esterno in qualsiasi altro modo». La proposta in questione sembra evitare attentamente di nominare i sistemi ASAT con base a terra del tipo che entrambe le superpotenze stanno sviluppando. Ove si eccettui la proposta Breznev del 1981, nulla che non fosse di natura formale è accaduto in questo campo dopo l'aggiornamento, nel 1979, dei negoziati ASAT.

Ultimamente si è parlato molto della possibilità di porre in orbita nello spazio «armi a energia diretta», con riferimento di solito a tre categorie di armi: i laser ottici convenzionali, gli ordigni a fasci di particelle e una nuova categoria di sistemi, indicati enigmaticamente dai loro fautori come «armi nucleari della terza generazione», che come è noto comprendono laser a raggi X alimentati da esplosioni nucleari. Queste tre categorie molto differenti si trovano in una fase molto diversa di ricerca e di sviluppo. La prima categoria, i laser ottici, è di gran lunga la più avanzata. Un sistema di questo genere non è stato ancora progettato in tutti i particolari, ma presumibilmente dovrebbe comprendere vari sottosistemi, alcuni dei quali sarebbero simili (sia pure con dimensioni e capacità maggiori) a ordigni che già sono stati fatti funzionare in laboratorio sulla Terra.

Un sistema di armi basate sul laser in grado di distruggere bersagli «cooperativi» potrebbe probabilmente essere spiegato nello spazio, sia pure a un costo elevatissimo, prima della fine del secolo. (Mi sto riferendo qui ai sistemi ad alta energia che distruggono materialmente il loro bersaglio, non a sistemi a bassa potenza che accecano i sensori ottici o causano altrimenti danni speciali limitati.) Le altre due categorie di armi a energia diretta sono ancora soltanto nella fase di ricerca. Nessuno sviluppo dell'uno o dell'altro sistema è stato ancora portato a termine o è attualmente in progetto, e a mio avviso non c'è la benché minima possibilità che un ordigno operativo di uso pratico, sia pure di capacità limitata, venga spiegato nello spazio nel corso di questo secolo.

Sebbene non sia attualmente proibito da nessun accordo internazionale, lo sviluppo di sistemi di questo genere è fortemente ostacolato da un fattore ancora più potente: il costo estremamente elevato. Le migliori stime disponibili indicano che il solo costo dello sviluppo di un «posto di combattimento laser» sia superiore di dieci volte a quello di analoghi sistemi moderni ad alta tecnologia (ciò significa decine di miliardi di dollari anziché miliardi di dollari) e che lo spiegamento di un sistema capace di intercettare più di una piccola frazione della forza strategica sovietica verrebbe a costare una grossa frazione di un bilione di dollari. Questi costi enormi hanno impressionato, almeno fino a questo momento, i funzionari responsabili e, nonostante le iniziative da «guerre stellari» del presidente Reagan, non esiste ancora nessun piano serio di posti di combattimento di questo genere basati su uno qualunque di vari tipi ipotetici di «raggi della morte».

Entrambi, Stati Uniti e Unione Sovietica, hanno caratteristiche uniche che rendono ognuna delle due superpotenze un partner difficile di un negoziato, sia quando trattano fra di loro sia quando trattano con terzi. Nel caso degli Stati Uniti la più seria di queste difficoltà particolari è dovuta a certi meccanismi fondamentali del sistema americano di governo: la necessità di una maggioranza di due terzi al Senato per la ratifica di un trattato e la campagna presidenziale eccezionalmente lunga che il paese effettua ogni quattro anni.

Fu l'esigenza di una maggioranza di due terzi al Senato a tenere gli Stati Uniti fuori della Società delle Nazioni dopo la prima guerra mondiale, a impedire l'adesione al Protocollo di Ginevra sulla guerra chimica nel 1925 finché non furono passati 50 anni buoni e, più di recente, a impedire di ratificare il Threshold Test Ban Treaty, il Peaceful Nuclear Explosions Treaty e il SALT 2 Treaty. Probabilmente questi trattati sarebbero stati ratificati, se fosse bastato un voto a maggioranza semplice. A dire il vero, questa situazione particolarmente difficile e imbarazzante è proprio lo scopo della norma. L'idea dei padri fondatori del paese era che gli Stati Uniti dovessero evitare le «complicazioni straniere». Quando il paese era veramente isolato da due grandi oceani, la cosa probabilmente aveva un certo senso. Ma in un mondo che è strettamente integrato dall'alta tecnologia e in cui i missili sovietici in Siberia sono a mezz'ora di distanza dai loro obiettivi negli Stati Uniti, la questione diventa drasticamente diversa.

Il problema non è costituito semplicemente dalla mancata ratifica di certi trattati. Ancor più importante, a mio parere, è l'effetto che questo fattore ha sull'andamento delle trattative. All'inizio del proprio mandato ogni presidente scopre che il controllo degli armamenti è ancor più controverso di quanto pensasse e che la misura ultima dell'accettabilità di ciò che egli sta negoziando è quella di riuscire ad avere una maggioranza di due terzi al Senato. Di conseguenza la politica relativa al controllo degli armamenti viene elaborata pensando, non al consenso nazionale né alla maggioranza del Congresso, ma piuttosto con l'obiettivo di ottenere in un modo o nell'altro l'appoggio di quei cinque o sei senatori che metterebbero sulla strada giusta per raggiungere la maggioranza richiesta. Nel bene e nel male questo fattore ha influenzato in grande misura la politica di ogni presidente, da Kennedy a Carter, e le istruzioni negoziali impartite alle delegazioni oltremare sono sempre state caratterizzate da estrema cautela e da estremo conservatorismo. Al contrario di quanto comunemente si crede, questo significa spesso che, via via che in un negoziato il tempo passa, la posizione americana subisce repentini mutamenti, non in risposta a ciò che può essere stato proposto dall'altra parte, ma in una difficile e a volte futile manovra presidenziale intesa ad accontentare alcuni senatori chiave e quei membri dell'esecutivo, soprattutto fra i militari in divisa, che hanno particolare influenza su quei senatori.

Grave è anche il problema costituito dalla campagna presidenziale quadriennale. Poiché la politica del controllo degli armamenti è intrinsecamente controversa, tutti i nuovi presidenti non tardano molto a scoprire che è assolutamente indispensabile consultare tutti gli elementi interessati dell'esecutivo prima di poter stabilire finalmente una politica di controllo degli armamenti ed elaborare istruzioni particolareggiate. Questo processo prende tipicamente quasi tutto o tutto il primo anno del mandato di un presidente appena eletto. Più tardi, nel quarto anno del suo mandato, il presidente si trova di fronte alla campagna per la nomina dei candidati all'interno del proprio partito, seguita dalla campagna generale nella quale deve affrontare il candidato del partito avversario.

I problemi creati dalla campagna presidenziale americana eccezionalmente lunga appaiono in piena evidenza quando si considerano i problemi che prima il presidente Johnson e poi il vicepresidente Humphrey ebbero con questo tipo di questione nel 1968, i problemi che il presidente Ford ebbe a proposito dei negoziati SALT 2 e di quelli sul Canale di Panama durante la sua campagna all'interno del partito contro Ronald Reagan nel 1976, e i problemi che il presidente Carter ebbe prima con il senatore Kennedy e poi con Ronald Reagan nel 1980 (questi ultimi esacerbati ovviamente in misura notevole dagli avvenimenti dell'Iran e dell'Afghanistan). Insomma, gli Stati Uniti, almeno in questa era di presidenze quadriennali, sono in grado di negoziare seriamente soltanto per due anni ogni quattro e anche allora in un modo che costringe il presidente ad accentrare particolare attenzione su quelle poche persone che possono far spostare verso i due terzi la maggioranza dei favorevoli.

Oltre a questi problemi peculiarmente statunitensi, anche le difficoltà burocratiche che affliggono tutte le società complesse hanno una forte influenza negativa sulla capacità degli Stati Uniti di negoziare. Questo è dovuto soprattutto al fatto che, per raggiungere il proprio obiettivo, chi vota contro deve soltanto bloccare o rallentare in qualche modo il procedimento, mentre chi vuole aiutare il presidente a raggiungere il suo obiettivo riguardo al controllo degli armamenti deve trovare proprio la strada giusta su un territorio spesso inesplorato.

Forse i problemi più gravi con i sovietici come partner di un negoziato sono costituiti dalla loro propensione alla segretezza e dal loro sistema politico rigorosamente controllato. Questi due problemi insieme sono alla radice delle difficoltà che gli americani incontrano sempre quando trattano il problema del controllo. Gli Stati Uniti hanno «mezzi tecnici nazionali» di controllo straordinariamente potenti e raffinati, fra cui satelliti per le osservazioni, stazioni sismiche in tutto il mondo e altri sistemi di spionaggio. Con questi mezzi possono e vengono effettivamente a sapere molte cose su quello che avviene nell'Unione Sovietica, e quasi tutti coloro che conoscono bene questi mezzi tecnici e il loro rendimento affermano che essi sono sufficienti a verificare il rispetto di tutti i trattati che sono stati firmati fino a questo momento. Ciononostante, è necessario e prudente che il Governo americano, e soprattutto i vari servizi di spionaggio, classifichino i particolari di questi sistemi e tengano segrete le valutazioni del loro lavoro.

Data l'effettiva necessità di segretezza da parte degli Stati Uniti, e dato che nell'ambito del sistema c'è sempre qualcuno che non è pienamente soddisfatto delle prestazioni dei mezzi tecnici nazionali di controllo, non dovrebbe destare sorpresa il fatto che quei funzionari e altri membri dello Stato americano che hanno opinioni estremamente negative sull'Unione Sovietica, in generale, e che adottano un'impostazione cauta del problema del controllo degli armamenti, in particolare, siano dubbiosi sulla capacità dei mezzi tecnici di scoprire ogni sotterfugio che possa essere escogitato. L'Unione Sovietica, in fin dei conti, è un paese enorme con grandi aree interamente vietate agli stranieri, e non è semplice paranoia supporre che in tali aree possa venir fatto qualcosa d'importante, ma di cui i servizi di spionaggio americani non siano a conoscenza.

Il controllo politico che nell'Unione Sovietica si estende a tutti i livelli della vita aumenta molto le difficoltà del problema. È banale, ma vero che non vi sono partiti d'opposizione organizzati, che non vi sono gruppi d'interesse pubblico, né gruppi di pressione o di informazione come quelli che in quasi tutte le altre società industriali sono capaci di scoprire e rendere di pubblica ragione i misfatti governativi, ivi comprese le violazioni dei trattati. Nella mia discussione sulla messa al bando degli esperimenti ho accennato al fatto che uno dei problemi generali che hanno continuato a esistere in tutto il periodo post-bellico di negoziati bilaterali è che gli americani chiedono sempre mezzi di controllo più intrusivi di quanto i sovietici siano disposti anche a discutere e tanto meno ad accettare, mentre i sovietici continuano a protestare che gli americani cercano sempre di spiarli, di minare alla base la loro sovranità e di interferire altrimenti nei loro affari interni.

A mio parere questo problema continuerà a esistere in futuro a tempo indeterminato, e importanti elementi dello Stato americano continueranno a nutrire gravi dubbi sull'adeguatezza di qualsiasi sistema di controllo possibile e immaginabile. Personalmente non vedo nessun modo facile per superare queste difficoltà fintantoché, e a meno che, non si abbiano notevoli cambiamenti nel comportamento politico dell'Unione Sovietica, e non vedo nessuna ragione per attendersi una cosa del genere nel futuro prevedibile.

La situazione non è del tutto grigia. Le opinioni dei leader sovietici in fatto di segretezza e di sovranità si sono andate evolvendo, e certi loro vecchi atteggiamenti su questioni attinenti sono diventati meno rigidi. Per fare un esempio, il fatto che abbiano accettato la ricognizione mediante satelliti artificiali, che abbiano aderito a un accordo di massima sullo scambio di dati geofisici in connessione con il Threshold Test Ban Treaty, che abbiano acconsentito ad accettare speciali attività *in loco* da parte di osservatori stranieri nel Peaceful Nuclear Explosions Treaty, apposite stazioni sismiche e un sistema attentamente vincolato di ispezioni volontarie *in loco* nei negoziati per la messa al bando totale degli esperimenti sono tutti elementi a favore di questa evoluzione. Se poi questa evoluzione proceda in modo da fugare i sospetti americani è cosa che resta ancora da vedere.

La maturazione dell'RNA

Il DNA viene trascritto in RNA messaggero, poi tradotto in proteina. Nelle cellule dotate di nucleo, tra cui le cellule umane, molti eventi si svolgono nell'RNA nel periodo tra la trascrizione e la traduzione

di James E. Darnell, Jr.

Uno dei problemi fondamentali della biologia molecolare è quello di scoprire in che modo due cellule con la stessa dotazione di geni possano, ciascuna, produrre una differente serie di proteine. Con ogni probabilità, tranne poche eccezioni, tutte le cellule di un organismo pluricellulare hanno la stessa informazione genetica, codificata nelle sequenze di basi nucleotidiche che costituiscono il loro DNA. La sequenza nucleotidica nel filamento di DNA funge da insieme di istruzioni per l'assemblaggio delle proteine che dirigono tutte le funzioni cellulari. Si potrebbe pensare, pertanto, che tutte le cellule di un organismo pluricellulare sintetizzino le stesse proteine. Nella realtà, invece, ogni tipo di cellula produce solo una parte caratteristica di tutte le proteine codificate nel suo DNA. Come viene controllato, allora, l'apparato genetico di ogni tipo di cellula, perché produca solo una serie specifica di proteine e nelle giuste quantità?

Le cellule degli organismi pluricellulari sono eucarioti, cioè hanno un nucleo ben definito. In esse, la sequenza dei nucleotidi all'interno di un gene, cioè il segmento di DNA che codifica per una sola proteina, viene trascritto in una molecola di RNA, che fuoriesce dal nucleo e passa nel citoplasma circostante, dove viene tradotta in proteina. Così, nella cellula eucariote, sono possibili parecchi livelli di regolazione genetica. Potrebbe essere esercitato un controllo per individuare quali segmenti del DNA vengono trascritti in RNA, quali sequenze trascritte vengono trasportate nel citoplasma e la velocità alla quale viene tradotto un particolare RNA. Da tempo si è dato per scontato che il controllo della trascrizione e il controllo della traduzione svolgano una funzione nella regolazione dei geni degli eucarioti. Tuttavia, ancora nel 1975 la conoscenza di ciò che avviene nell'RNA tra la trascrizione e la traduzione era lacunosa.

Alla fine degli anni settanta, divenne evidente che sono necessarie parecchie operazioni significative per trasformare il prodotto di trascrizione del DNA, l'RNA primario, o appena formato, nell'RNA messaggero (m-RNA) maturo, necessario per la traduzione. Alle due estremità del filamento di RNA vengono aggiunte differenti strutture. Certi nucleotidi lungo la catena subiscono una modificazione chimica. Tuttavia, è stata del massimo interesse la scoperta che il prodotto di trascrizione, l'RNA primario, viene frequentemente tagliato e nuovamente saldato per dare un m-RNA più corto. In alcuni casi, questo RNA può essere tagliato e nuovamente saldato in più di un modo e dà origine così a differenti m-RNA e quindi a differenti proteine. Questo taglio e successiva saldatura differenziali potrebbero anche essere una forma di controllo genetico e, in alcuni casi, è stato dimostrato che è proprio così. È probabile che il processo di maturazione dell'RNA, successivo alla trascrizione, sia necessario per molti m-RNA, ma che sia una forma di regolazione solo per pochi. La sua scoperta ha nondimeno aggiunto una nuova dimensione al quadro dell'espressione genica, il che è in sé un passo avanti significativo. Inoltre ha permesso di formulare interessanti ipotesi sulla primissima storia dei geni e delle cellule.

La trascrizione del DNA

Come il DNA, l'RNA è una catena polimera, le cui unità monomere sono nucleotidi. Nell'RNA i quattro nucleotidi sono quelli dell'adenina, della citosina, della guanina e dell'uracile, le cui abbreviazioni sono *A*, *C*, *G* e *U*. (Nel DNA il posto di *U* è occupato dalla timina, *T*.). Ogni nucleotide consta di una base azotata e di uno zucchero a cinque atomi di carbonio. Questi cinque atomi sono in genere designati con numeri. Ogni coppia di nucleotidi adiacenti è unita da un gruppo fosfato che forma un legame tra il carbonio 5' del residuo di zucchero di un nucleotide e il carbonio 3' del residuo di zucchero del nucleotide adiacente. Così la molecola di RNA ha una sua direzionalità: un'estremità è la terminazione 5' e l'altra la terminazione 3'.

La trascrizione del DNA in RNA viene realizzata da una RNA-polimerasi, uno dei tre tipi di polimerasi presenti in una cellula eucariote. L'enzima si lega al DNA in corrispondenza di un giusto sito «di inizio» per l'RNA e sceglie il primo nucleotide, che diventa l'estremità 5' del filamento di RNA. Quindi si sposta rapidamente lungo il filamento di DNA, aggiungendo i giusti nucleotidi al filamento di RNA. L'RNA poi, appena formato, si separa dal DNA: per prima si stacca l'estremità 5' e il complesso della molecola non si allontana fino a quando non si è formata anche l'estremità 3'.

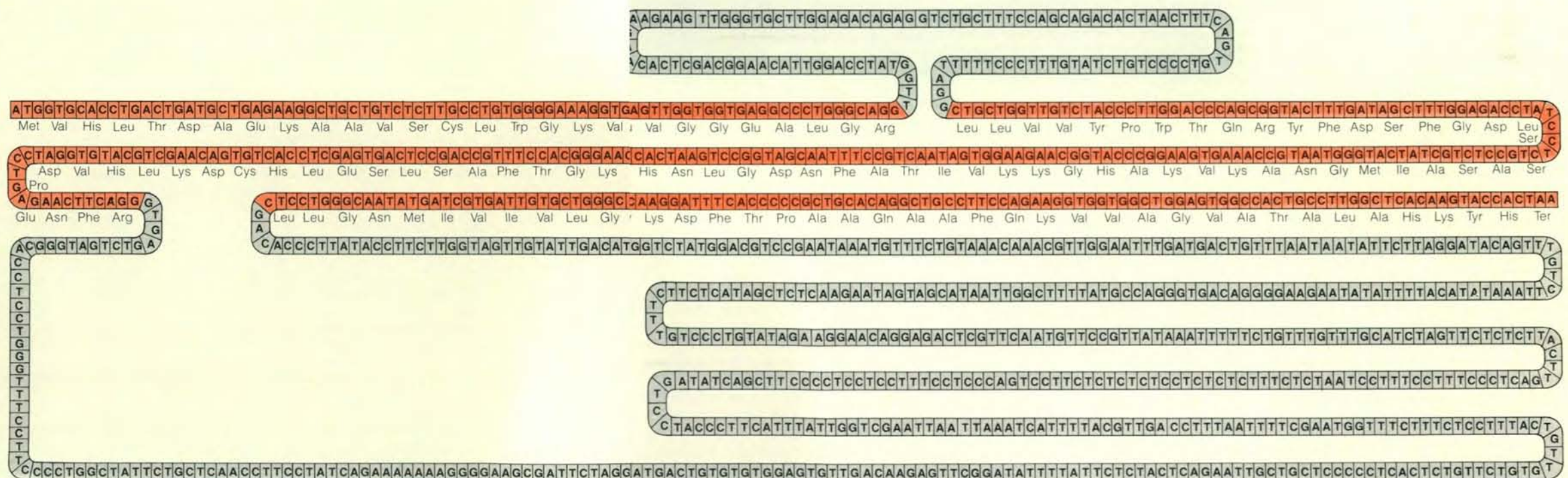
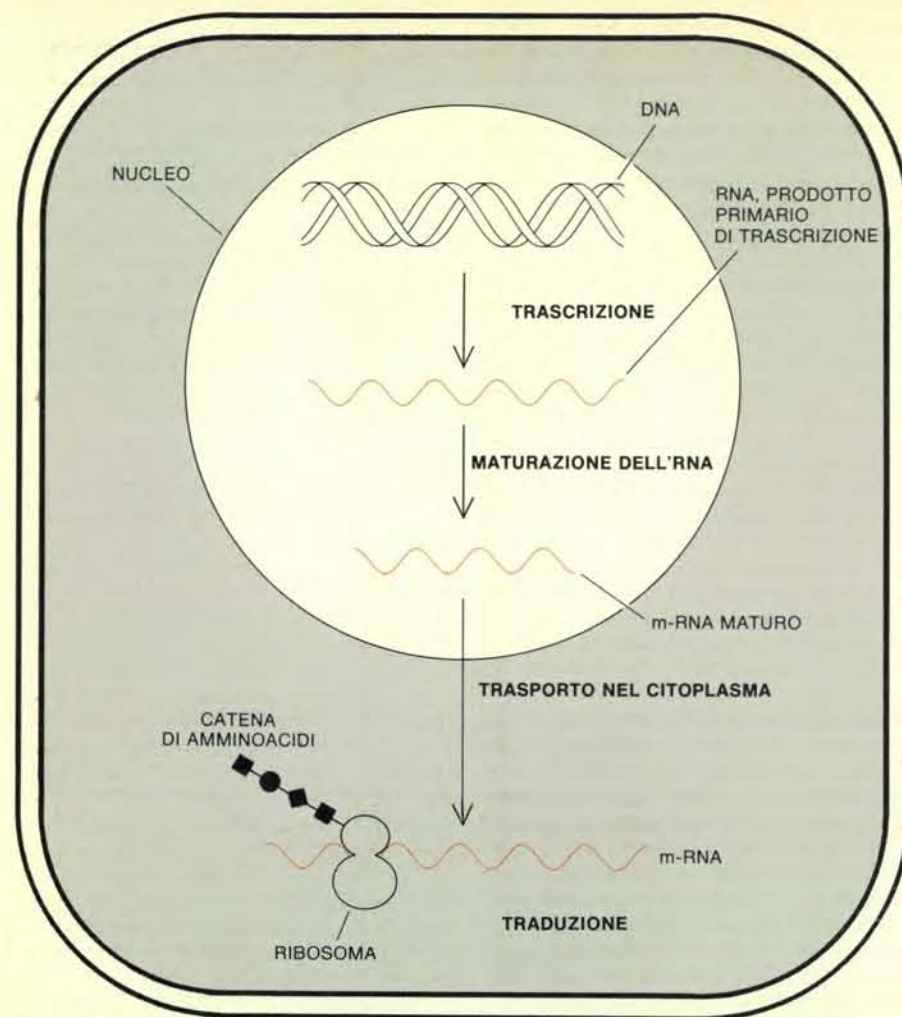
Quando la RNA-polimerasi copia un tratto di filamento del DNA, il nucleotide che viene aggiunto all'RNA è complementare a quello presente sul DNA. A causa della struttura molecolare dei nucleotidi, ciascuno di essi può formare legami a idrogeno con uno solo degli altri tre. Così *C* e *G* possono formare legami a idrogeno solo tra loro, e *A* e *U* (o *T*) lo stesso. Sono proprio questi legami a idrogeno tra basi (o meglio nucleotidi) complementari che fanno sì che i due filamenti di DNA formino una doppia elica. Analogamente, un segmento di RNA può formare con un segmento di DNA una molecola a duplice filamento se le basi sui due filamenti sono complementari.

Pertanto, l'RNA trascritto è complementare al filamento di DNA da cui proviene. Poiché conserva l'informazione di quella parte del genoma (l'intero complemento d'informazione genetica), esso può incorporare il codice per la catena di unità amminoacidiche che costituisce una proteina. Il codice viene letto sull'RNA a tre nucleotidi alla volta e ogni tripletta di nucleotidi, che codifica per uno dei 20 diversi amminoacidi, è chiamata codone. Nel processo di traduzione, la lettura del codone porta all'aggiunta del corrispondente amminoacido alla molecola proteica che si sta formando.

I tre RNA

L'RNA messaggero, l'RNA che ha una funzione codificatrice nei riguardi delle proteine, non è il solo tipo di RNA. Ve ne sono perlomeno due altri tipi principali, ognuno dei quali ha una funzione significativa nella sintesi proteica. L'RNA ribosomale (r-RNA) è un componente del ribosoma, l'organello citoplasmatico che funge da piattaforma per la traduzione. Il ribosoma consta di due subunità, designate come grande e piccola: ogni subunità incorpora una molecola di r-RNA e tra 25 e 50 molecole proteiche. L'RNA di trasporto (t-RNA) funziona come un uncino molecolare per attaccare l'amminoacido successivo alla fine di un filamento proteico; ciascuno dei diversi t-RNA riconosce un particolare amminoacido.

Sul ribosoma, l'RNA di trasporto viene a trovarsi assieme all'RNA messaggero e agli enzimi, che servono a connettere il giusto amminoacido alla molecola proteica in via di formazione e a far progredire l'intero apparato per leggere il codone successivo. Così, nella sintesi di una proteina, il t-RNA e l'r-RNA hanno ruoli ripetitivi che sono realizzati seguendo le istruzioni dell'm-RNA. Nelle cellule umane, vi è all'incirca un milione di copie di ogni tipo di t-RNA assieme a circa cinque milioni di copie di ciascuno dei due tipi principali di r-RNA: una per ogni ribosoma presente nella cellula.

A causa della loro abbondanza, gli RNA di trasporto e gli RNA ribosomali sono più facili da studiare degli RNA messaggeri ed erano già stati intensivamente analizzati prima che si fosse raggiunta una buona conoscenza anche dell'm-RNA. Il primo esempio di maturazione dell'RNA è stato trovato proprio nell'RNA ribosomale. Nel 1961, Klaus Scherrer e Harriet Latham hanno collaborato con me a una ricerca condotta in un nuovo laboratorio per lo studio della cellula animale e dei virus animali presso il Massachusetts Institute of Technology. Abbiamo cominciato ad analizzare l'RNA di cellule umane, sviluppatesi su un terreno colturale. Volevamo, in particolare, confrontare gli RNA appena

Un gene che codifica per una proteina è una sequenza di nucleotidi (*A, C, G, T*) nel DNA. La sequenza riportata qui si riferisce al gene per la globina beta nel topo. La globina beta è una delle due catene amminoacidiche dell'emoglobina. La sequenza effettivamente codificante del DNA è interrotta da due «introni» (*in grigio*). I tre segmenti di sequenza codificante sono «esoni» (*in colore*). Gli amminoacidi della globina beta sono indicati con le rispettive sigle vicino ai corrispondenti nucleotidi. Perché il messaggio genetico possa essere tradotto nella catena amminoacidica della proteina, l'RNA deve venir trascritto dal DNA. Le sequenze dette introni devono quindi essere rimosse in modo che l'RNA messaggero (m-RNA) abbia la giusta sequenza codificante. Gli introni sono caratteristici dei geni delle cellule eucarioti, che, diversamente da quelle batteriche, hanno un nucleo ben definito. I nucleotidi *ATG* all'inizio della sequenza (*in alto a sinistra*) sono il segnale di inizio della sintesi proteica. I nucleotidi *TAA* alla fine sono il segnale per la fine della sintesi proteica. La sequenza è stata definita da Philip Leder e collaboratori alla Harvard Medical School.

La regolazione genetica in una cellula eucariote avviene a parecchi livelli. Nel nucleo il DNA viene dapprima trascritto in un RNA primario, che è una copia complementare di un segmento di DNA. Questo prodotto di trascrizione primario viene quindi trasformato in m-RNA maturo: cioè l'RNA messaggero al quale sono state asportate le sequenze non codificanti. L'm-RNA è poi trasportato dal nucleo al citoplasma, dove viene tradotto dai ribosomi nella catena amminoacidica di una proteina. L'espressione dei geni può essere controllata durante la trascrizione e la maturazione nel nucleo, durante la traduzione nel citoplasma e durante il trasporto dal nucleo al citoplasma.

trascritti nella cellula con i tipi di RNA stabile presenti in grandi quantità: i t-RNA e i due tipi di r-RNA.

I batteri sono procarioti, cioè le loro cellule non hanno un nucleo ben definito. In esse la trascrizione e la traduzione si svolgono simultaneamente: l'RNA messaggero comincia a essere tradotto mentre è ancora in fase di trascrizione sulla molecola di DNA. Come le cellule eucarioti, tuttavia, le cellule batteriche dipendono dai ribosomi e dai t-RNA; pertanto, buona parte dei primi lavori sui tre tipi stabili di RNA fu compiuta con cellule di questo genere.

I tre tipi di RNA stabile e abbondante possono venir distinti sulla base delle loro dimensioni. Tutti i t-RNA sono piccoli: sono lunghi all'incirca 80 nucleotidi. I due principali tipi di RNA ribosomale sono, invece, sostanzialmente più grossi. La dimensione di una molecola di RNA si misura spesso in base alla velocità con cui la molecola precipita quando viene centrifugata a velocità elevata in un'ultracentrifuga. Questo metodo si chiama analisi della sedimentazione zonale e la velocità di sedimentazione si esprime in unità Svedberg, abbreviate con *S*. Quanto più lungo è l'RNA, tanto più elevato è il numero di *S*. Le due forme principali di r-RNA nelle cellule eucarioti vengono designate come 18*S* e 28*S*.

Nei nostri primi esperimenti, Scherrer, Latham e io abbiamo cercato di trovare se i prodotti iniziali, nella trascrizione dell'RNA, hanno le stesse dimensioni dei tipi abbondanti di RNA o, alternativamente, se l'RNA viene trascritto in unità che sono più lunghe o più corte delle molecole abbondanti. I nucleosidi, cioè i precursori dei nucleotidi, possono essere marcati con vari atomi radioattivi. Se alle cellule in accrescimento sono forniti nucleosidi radioattivi, le subunità marcate entreranno nei pool di nucleotidi che forniscono il materiale costruttivo per i filamenti di RNA. Le molecole radioattive saranno quindi aggiunte a nuovi filamenti di RNA per opera delle RNA-polimerasi. La marcatura offre un mezzo per identificare l'RNA appena formato: dopo che gli RNA marcati sono stati estratti dalla cellula, possono venir separati in base alla loro dimensione, analizzandone la sedimentazione, e confrontati con RNA non marcato, già nella cellula in precedenza.

La maturazione dell'r-RNA e del t-RNA

Inizialmente le cellule hanno ricevuto i nucleosidi marcati per soli cinque minuti. Dopodiché, la radioattività non è comparsa nell'RNA con le dimensioni corrispondenti ai tre preesistenti tipi stabili. Essa è comparsa, invece, nel nucleo delle cellule in un gruppo assortito di molecole con un'ampia gamma di dimensioni. Molte molecole marcate erano parecchio più lunghe anche del più lungo RNA ribosomale, che consta di 5000 nucleotidi. Come vedremo, tali grosse molecole, che si marcavano rapidamente, hanno dimostrato di avere un notevole significato nell'attività dell'RNA messaggero. I prodotti di trascrizione che erano rimasti marcati entro i cinque minuti erano di tante dimensioni diverse da non poter essere inclusi in poche categorie sulla base della loro dimensione. Quando il tempo durante il quale i nucleosidi marcati venivano forniti alla cellula fu aumentato da cinque a 60 minuti, gruppi distinti di molecole marcate e dotate di differenti dimensioni risultarono, invece, evidenti. Una delle molecole più grosse, che comparve dopo 25-30 minuti di marcatura, aveva un valore di sedimentazione di 45*S*.

La composizione in basi dell'RNA 45*S* (cioè le quantità relative di *A*, *C*, *G* e *U*) era molto simile a quella degli RNA 18*S* e 28*S*. La molecola 45*S* risultò presto correlata con gli RNA ribosomali in un modo diverso ed estremamente interessante. La trascrizione può essere interrotta rapidamente per opera dell'antibiotico actinomicina, che si lega strettamente al DNA. Abbiamo fornito l'antibiotico alle cellule dopo che il marcatore era già stato assorbito dalla molecola 45*S*. A questo punto dal DNA non veniva più trascritto RNA; pertanto il marcatore si trovava solo in quelle molecole che erano state trascritte in precedenza. Poco dopo la somministrazione dell'actinomicina esso scompariva dalle molecole 45*S* e ricompariva negli RNA 18*S* e 28*S*.

Abbiamo avanzato l'ipotesi che gli RNA 45*S* siano un lungo prodotto primario di trascrizione che viene tagliato per dare i due RNA ribosomali finiti. Da allora essa è stata convalidata da molti esperimenti. Metodi perfezionati di definizione delle sequenze nucleotidiche hanno reso possibile la determinazione delle sequenze di basi in ambedue i geni per l'RNA ribosomale e per gli r-RNA stessi. È evidente oggi che in molti tipi di cellule eucarioti ivi incluse quelle di lievito, delle piante, delle rane e dei topi, un lungo precursore dell'r-RNA viene tagliato e fornisce molecole finite più corte. Esperimenti recenti hanno mostrato che questo accade perfino nei batteri.

Alla fine degli anni sessanta, inoltre, Roy H. Burdon dell'Università di Glasgow e Deborah Bernhardt nel mio laboratorio al MIT hanno trovato che anche l'RNA di trasporto si forma in seguito a un processo di maturazione. Nel t-RNA, però, la differenza di lunghezza tra prodotto di trascrizione primario e molecola finita è solo tra 20 e 30 basi: un precursore di circa 100 nucleotidi viene accorciato e dà il prodotto finito che consta di 80 nucleotidi.

All'inizio degli anni settanta, in molti tipi di cellule il processo di maturazione dell'RNA ribosomale e dell'RNA di trasporto era stato ormai stabilito in maniera definitiva. Si delineò a quel punto la questione se anche l'RNA messaggero fosse fatto maturare in modo analogo. Sembrava improbabile che esso lo fosse nelle cellule procarioti, visto che nei procarioti i ribosomi traducono l'RNA messaggero mentre è ancora in fase di trascrizione. Negli eucarioti, d'altra parte, dove la trascrizione e la traduzione sono separate nel tempo e nello spazio, la maturazione dell'm-RNA era un evento ben distinto e interessante.

Gli RNA nel poliribosoma

Molte più cose andavano tuttavia trovate a proposito dell'RNA messaggero prima che il problema della sua maturazione potesse essere affrontato direttamente. Quando, negli anni sessanta, venne scoperto il processo di maturazione dell'RNA ribosomale, l'RNA messaggero delle cellule eucarioti non era stato ancora isolato. Era già stato dimostrato però che gli amminoacidi sono congiunti in proteine nei poliribosomi (o polisomi), grappoli di ribosomi presenti nel citoplasma. Si presunse, dunque, che l'RNA messaggero fosse presente in essi, ma senza ancora disporre di prove.

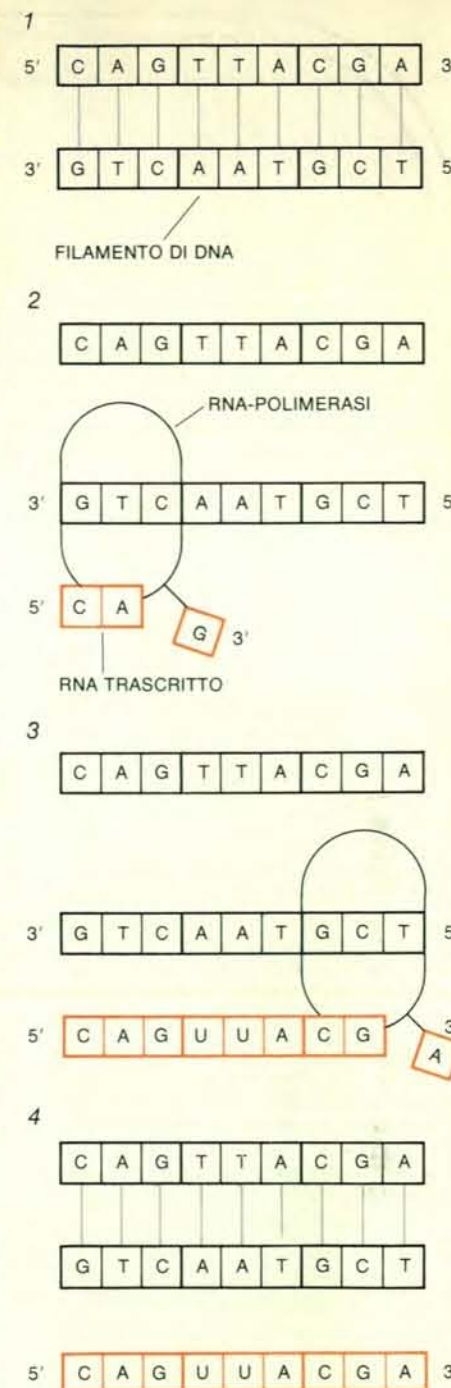
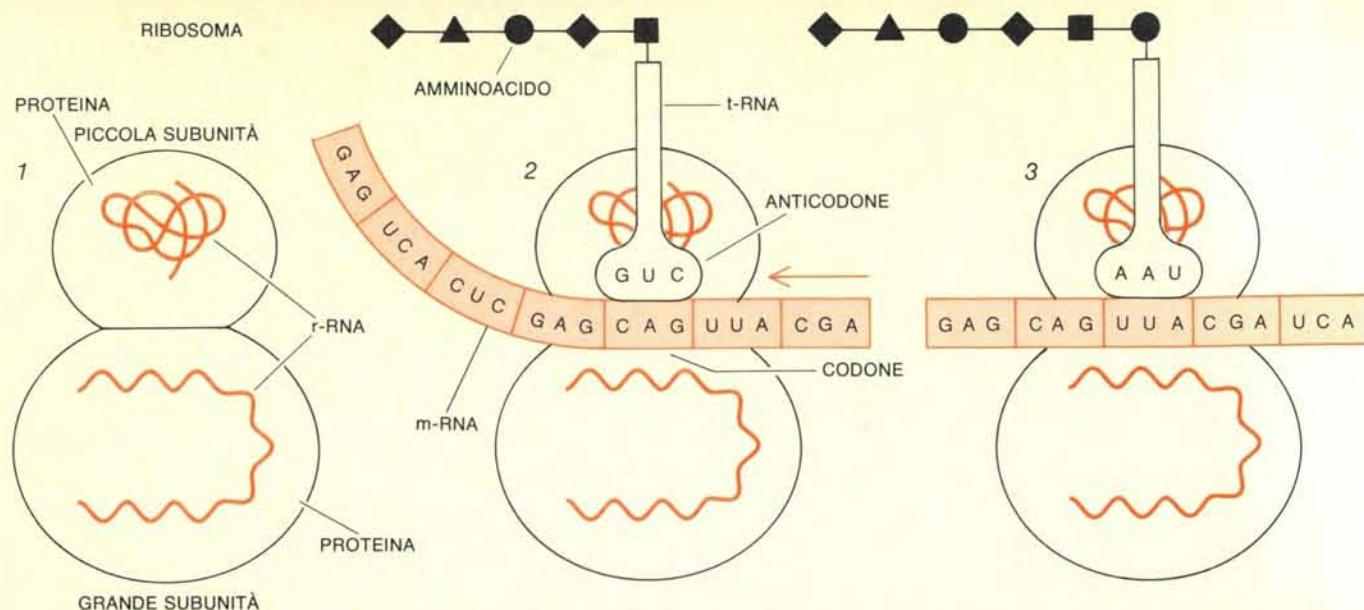
In un tentativo di trovare l'ipotetico messaggero, Sheldon Penman, Yechiel Becker e Scherrer nel mio laboratorio al MIT hanno coltivato le cellule tumorali umane, chiamate cellule HeLa, in presenza di nucleosidi marcati, i quali venivano incorporati nell'RNA; in seguito, questo RNA marcato era estratto dai poliribosomi. Una parte dell'RNA estratto era naturalmente RNA ribosomale, ma questo dopo una breve esposizione ai precursori radioattivi non era rimasto marcato. Vi era, invece, un altro gruppo di molecole di RNA nei poliribosomi che era stato marcato rapidamente: queste molecole avevano una composizione in basi diversa da quella dell'r-RNA. Gli RNA poliribosomali, come vennero chiamati, avevano una lunghezza variabile tra 500 e 3000 nucleotidi. Abbiamo avanzato l'ipotesi che ciascun RNA poliribosomale includesse una varietà di tutti gli RNA messaggeri presenti nella cellula.

La composizione in basi degli RNA poliribosomali è stata confrontata con quella delle lunghe molecole trovate nel nucleo durante la ricerca sul processo di maturazione dell'RNA ribosomale. L'insieme di RNA non ribosomali marcati rapidamente era stato denominato RNA nucleare eterogeneo (hn-RNA) e non fu difficile dimostrare che tale RNA aveva una composizione in basi molto simile a quella dell'RNA poliribosomale, brevemente marcato. Sembrava dunque che l'hn-RNA potesse essere un precursore della molecola di messaggero utilizzata nella traduzione che aveva luogo nei poliribosomi.

Il guaio con la suddetta ipotesi era che una parte degli hn-RNA aveva 5000 nucleotidi o più, mentre i più lunghi fra gli RNA poliribosomali avevano una estensione di soli 3000 nucleotidi. Tenendo presente l'esempio della maturazione del t-RNA e dell'r-RNA, abbiamo suggerito che l'hn-RNA venisse accorciato nel nucleo in un processo di maturazione, prima di essere trasportato nel citoplasma. Tuttavia, non fu possibile per alcuni anni provare questa possibilità in modo diretto. La ragione è che le cellule eucarioti sintetizzano migliaia di tipi di RNA messaggero e, dato che non era stato isolato alcun m-RNA singolo, non fu possibile osservare la sintesi di una molecola di m-RNA per dire se era stata o no sottoposta a un processo di maturazione.

Un modo per aggirare questa difficoltà consiste nel basarsi sulla semplicità degli organismi molto meno organizzati delle cellule, e cioè i virus. Pertanto ci siamo rivolti allo studio di cellule eucarioti infettate da virus. L'infezione di una cellula animale da parte di un virus è parsa un eccellente sistema sperimentale per studiare le proprietà biochimiche dell'RNA messaggero e, in ultima analisi, la sua sintesi. Il genoma di un virus, che è contenuto nella parte centrale della particella virale, è una molecola di DNA o di RNA. Quando il genoma virale entra nella cellula ospite, usurpa l'apparato sintetico della cellula in modo che essa sintetizzi gli acidi nucleici e le proteine virali. Alla fine del processo infettivo, nuove molecole virali si combinano nelle particelle virali che possono infettare altre cellule. In confronto alla cellula ospite di mammifero, il virus ha poche proteine e, pertanto, pochi RNA messaggeri. Questo fatto semplifica le cose. Dato che l'm-RNA virale deve essere tradotto dai ribosomi e dai t-RNA dell'ospite, esso può essere esaminato quasi come se fosse l'm-RNA di una cellula eucariote. Inoltre, alcuni virus a DNA penetrano nel nucleo cellulare, dove si svolge la sintesi enzimatica dell'm-RNA virale; pertanto gli stadi nella produzione di RNA possono anche essere studiati mediante un virus.

Il virus come strumento

La semplicità del virus venne sfruttata inizialmente per dimostrare che l'RNA poliribosomale che si marca in breve tempo è RNA messaggero. Penman e Becker nel mio laboratorio hanno trovato che nelle cellule infettate con il virus della poliomielite, che è un virus a RNA, questo RNA ha preso il posto nei ribosomi dell'RNA marcato in breve tempo. Parecchi ricercatori hanno quindi utilizzato un secondo metodo per dimostrare che l'm-

La trascrizione del DNA in RNA conserva l'informazione genetica codificata nel DNA. Ogni nucleotide di quest'ultimo può formare legami a idrogeno con un altro nucleotide soltanto, che viene detto il suo complementare: l'*A* (il nucleotide adenilico) si appaia con *T* (il nucleotide timilico) e *C* (il nucleotide citosilico) con *G* (il nucleotide guanilico). L'elica di DNA consta di due filamenti complementari tenuti uniti da tali legami (*1*). L'RNA è un filamento singolo costituito da nucleotidi e in cui *T* è sostituito da *U* (nucleotide uracilico). I filamenti sia di DNA sia di RNA sono direzionali: un'estremità viene indicata con 5' e l'altra con 3'. Un filamento di DNA viene trascritto in RNA dall'enzima RNA-polimerasi (*2*). Questo si sposta lungo un filamento del DNA, leggendo i nucleotidi uno alla volta e aggiungendo il nucleotide complementare a un filamento di RNA che si allunga dall'estremità 5' alla 3' (*3*). Quando quest'ultima si è completata, il prodotto di trascrizione si separa dal DNA (*4*).

La traduzione dell'RNA in proteina richiede due tipi di RNA oltre al messaggero che porta l'informazione genetica. Gli RNA ribosomali (r-RNA) sono componenti strutturali del ribosoma (*1*). Gli RNA di trasporto (t-RNA) fungono da uncino per aggiungere un nuovo amminoacido alla catena proteica in formazione; ogni t-RNA è specifico per un dato amminoacido. Ogni amminoacido nella proteina è specificato da un codone, sequenza di tre nucleotidi nel filamento dell'RNA messaggero. La molecola di t-RNA si lega all'm-RNA mediante un anticodone, gruppo di tre nucleotidi complementare al codone (*2*). Quando il ribosoma si sposta lungo la molecola del messaggero, l'anticodone riconosce il codone e forma con esso legami a idrogeno; l'amminoacido all'altra estremità del t-RNA si attacca alla catena proteica. Quando il ribosoma si sposta sul codone successivo, il t-RNA viene estromesso e sostituito da un altro t-RNA che corrisponde al nuovo codone (*3*).

-RNA virale penetra nei poliribosomi. Tale metodo è chiamato ibridazione molecolare e fu impiegato per la prima volta nei virus a DNA che infettano i batteri da Sol Spiegelman e Benjamin D. Hall dell'Università dell'Illinois. L'ibridazione molecolare si basa sul fatto che i filamenti di RNA e di DNA possono ibridarsi, ovvero formare una struttura a doppio filamento, solamente se le loro sequenze nucleotidiche sono complementari. Se si forma un ibrido, si può concludere che l'RNA, in origine, è stato trascritto a partire da quel DNA.

L'RNA estratto dai poliribosomi di cellule infettate da virus a DNA è stato mescolato con il DNA del virus e della cellula ospite. Fatto questo, l'RNA poliribosomale si è ibridato con il DNA del virus infettante, ma non con quello del mammifero ospite o di altri virus. Brian R. McAuslan della Princeton University ha compiuto un passo in più, mostrando che la struttura centrale del virus del vaiolo bovino, un grosso virus a DNA, può sintetizzare l'RNA da precursori nucleotidici. L'ibridazione molecolare venne impiegata per dimostrare che l'RNA sintetizzato dalla porzione centrale del virus era lo stesso dell'RNA poliribosomale presente nelle cellule infettate con il virus del vaiolo bovino. L'RNA poliribosomale era stato pertanto identificato in maniera incontrovertibile come RNA messaggero: si sapeva che era presente in corrispondenza del sito di traduzione ed era stato appaiato a filamenti complementari di DNA presenti nel genoma virale.

L'identificazione del messaggero ha costituito un decisivo progresso. Studiando gli m-RNA virali è stato possibile seguire la sintesi di specifiche molecole di messaggero per vedere quali operazioni sono eseguite sul prodotto di trascrizione. Sono stati rapidamente identificati parecchi tipi significativi di maturazione dell'RNA messaggero. Joseph R. Kates dell'Università del Colorado a Boulder ha notato che un segmento di ogni m-RNA sintetizzato dal virus del vaiolo bovino è ricco di nucleotidi dell'adenina (*A*). Mediante analisi chimica, Kates ha dimostrato che tale segmento si trova in corrispondenza dell'estremità 3' del messaggero. Poco dopo, quattro gruppi sotto la guida di Kates, di Mary Edmonds dell'Università di Pittsburgh, di George Brawerman della Tufts University e mia hanno dimostrato indipendentemente l'uno dall'altro che quasi tutti gli m-RNA poliribosomali di cellule animali hanno una sequenza di 150-200 nucleotidi adenilici *A* attaccati all'estremità 3'. Tale sequenza non è presente, invece, negli RNA di trasporto o ribosomali.

La «coda» di nucleotidi adenilici in corrispondenza dell'estremità 3' dell'RNA messaggero è stata denominata sequenza poli-*A* e il processo nel quale essa viene aggiunta è detto poliadenilazione. Si sa oggi che, nel nucleo cellulare, la coda è aggiunta all'hn-RNA (RNA nucleare eterogeneo) per opera di un enzima quasi subito dopo che l'hn-RNA è stato trascritto. Anche se la sequenza di eventi molecolari nella poliadenilazione non è stata ancora ricostruita in provetta, si è accertato che la RNA-polimerasi procede ben oltre il sito della sequenza poli-*A*: cioè la sede, sul DNA, che corrisponde all'estremità 3' dell'hn-RNA dove viene aggiunta la coda. Il filamento di RNA viene tagliato circa 20 nucleotidi dopo una sequenza segnale e parte di tale segnale è il gruppo *AAUAAA*, che è stato trovato in più di 100 m-RNA vegetali e animali. L'estremità 3' libera viene quindi allungata per aggiunta di circa 200 nucleotidi *A* e ciò grazie all'intervento di un enzima scoperto dalla Edmonds.

Teste e code

Quasi tutti gli RNA messaggeri che vengono trasportati verso il citoplasma hanno una coda poli-*A*. Tuttavia, un m-RNA non deve necessariamente avere una coda poli-*A* per essere tradotto. La funzione della coda sembra essere quella di proteggere l'm-RNA dalla demolizione operata dagli enzimi presenti nel citoplasma. Gli RNA messaggeri senza coda poli-*A* durano solo alcuni minuti nel citoplasma, mentre quelli con la coda possono essere stabili per ore o per giorni. Gli unici m-RNA noti che entrano nel citoplasma senza una coda poli-*A* sono quelli che vengono tradotti in quelle piccole proteine che si chiamano istoni e che si legano al DNA. Questi m-RNA per gli istoni, designati come poli-A^-, durano nel citoplasma meno di 30 minuti.

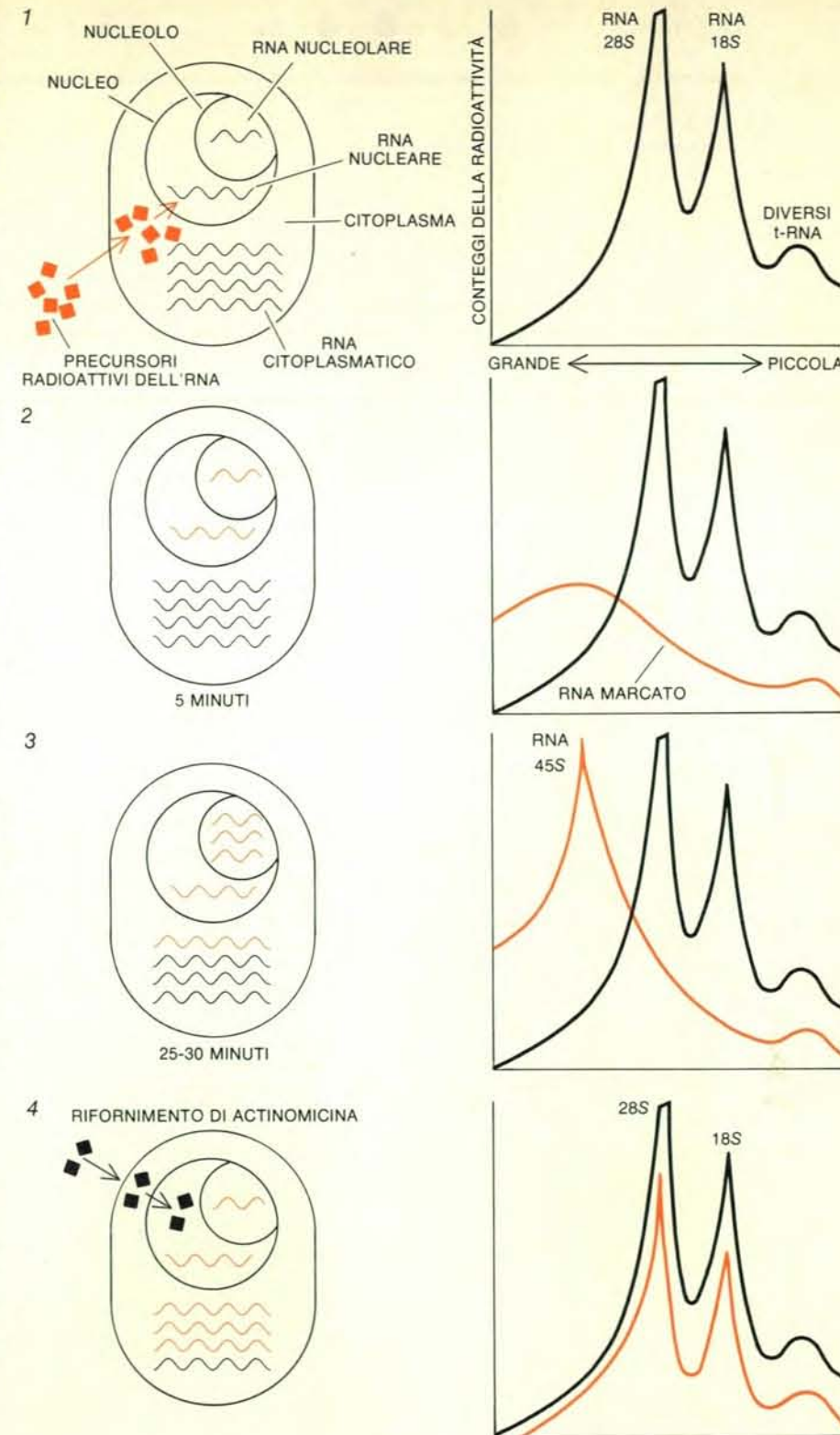
Oltre al suo interesse teorico, la sequenza poli-*A* ha avuto un notevole effetto pratico sulle ricerche relative al processo di maturazione dell'RNA messaggero. La ragione è che una molecola di m-RNA può essere afferrata per la sua coda poli-*A*. Ciò avviene sintetizzando un omopolimero polinucleotidico, cioè un filamento costituito da un solo tipo di nucleotide. Una colonna verticale di vetro viene riempita di fibre di cellulosa che hanno, aderente a esse, tale polimero; viene generalmente utilizzato un omopolimero consistente o di tutti *T* o di tutti *U*. Viene quindi fatto passare attraverso la colonna l'intero corredo di RNA della cellula. Le sequenze poli-*A* alla fine degli m-RNA possono ibridarsi con gli omopolimeri e rimanere nella colonna. L'RNA senza poli-*A*, principalmente RNA di trasporto e ribosomale, viene eliminato. Questa tecnica, chiamata cromatografia per affinità, è stata il primo mezzo pratico di isolamento in forma chimicamente pura dell'm-RNA di una cellula eucariote.

La disponibilità di m-RNA purificato ha dato un notevole impulso alle ricerche sulla maturazione dell'RNA. Sono stati in breve tempo trovati due altri procedimenti. Nel 1974 Robert P. Perry dell'Institute for Cancer Research ha scoperto che l'm-RNA della cellula animale non è composto solamente dei quattro nucleotidi regolari. Alcuni dei nucleotidi presenti nel filamento sono metilati, cioè presentano gruppi metilici supplementari (CH_3) che sono stati aggiunti a essi. Bernard Moss del National Institute of Allergy and Infectious Diseases, Aaron J. Shatkin del Roche Institute of Molecular Biology e Kin-Ichiro Miura dell'Istituto nazionale di genetica del Giappone hanno trovato che anche alcuni nucleotidi degli m-RNA virali sono metilati.

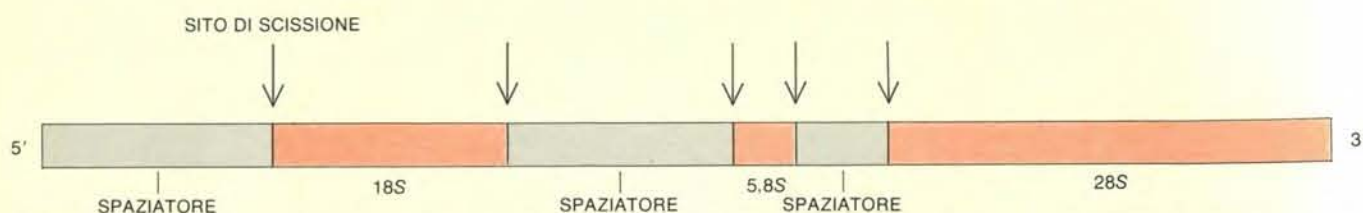
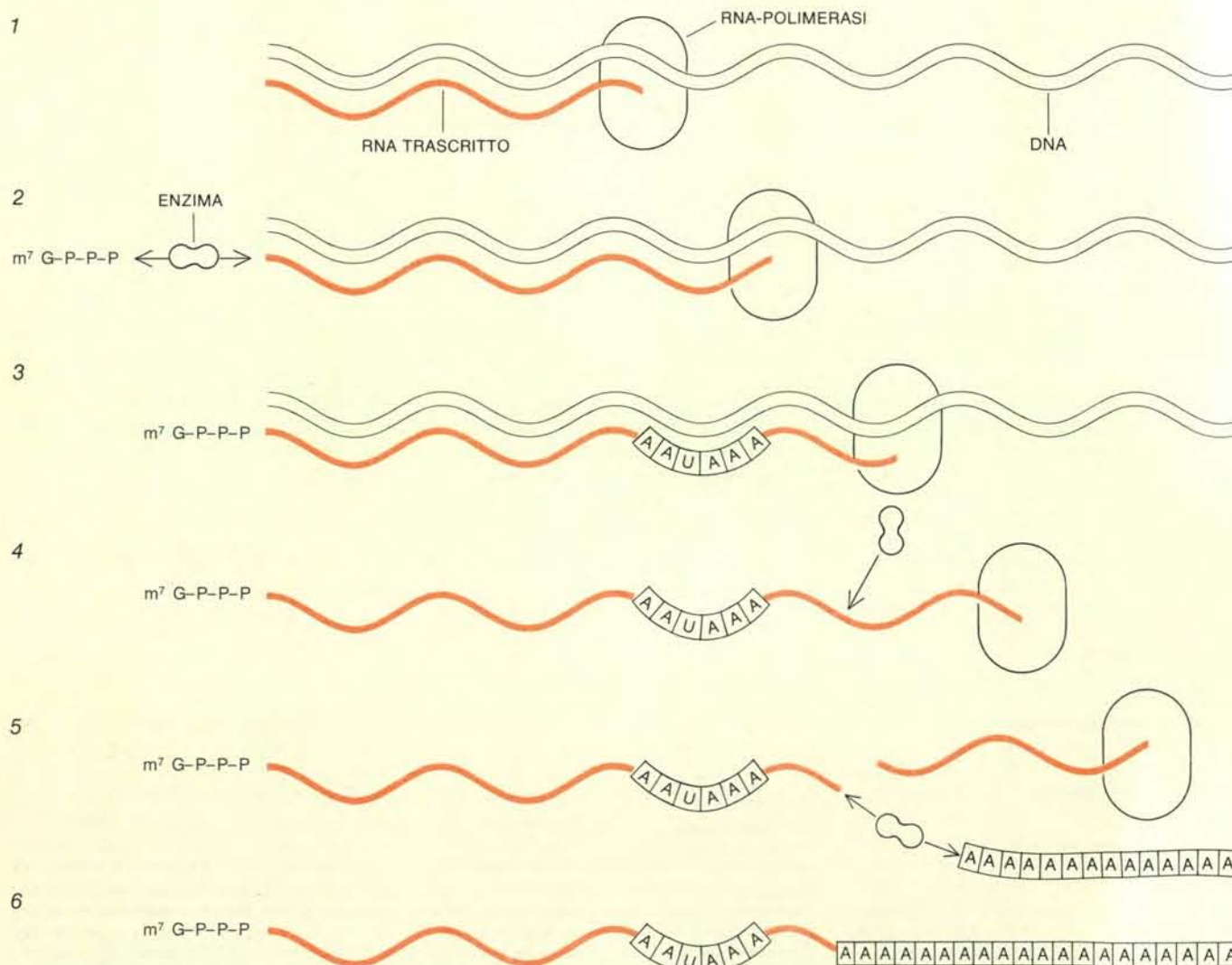
Gli enzimi chiamati ribonucleasi sono stati utilizzati per determinare dove si trovano, nell'RNA messaggero, le strutture metilate più significative. Le ribonucleasi scindono i legami tra nucleotidi adiacenti del filamento di RNA: ne risulta un insieme di singoli nucleotidi, ognuno con un gruppo fosfato attaccato. Invece, quando le ribonucleasi vennero utilizzate per l'RNA virale metilato, è stata trovata una struttura complessa che includeva parecchi gruppi fosfato, in aggiunta ai singoli nucleotidi.

Sulla struttura resistente alle ribonucleasi sono stati realizzati molti test chimici e oggi se ne conosce la costituzione. Tale struttura, detta «testa» in contrapposizione a «coda», e da alcuni «cappuccio» (*cap*), si trova all'estremità 5' di tutti gli m-RNA che sono stati osservati nelle cellule eucariote, a eccezione di un m-RNA di un unico virus. La componente più caratteristica della testa è un nucleotide guanilico (*G*), al quale è stato aggiunto un gruppo metilico. Il nucleotide guanilico viene unito all'estremità 5' dell'RNA mediante un legame che include tre gruppi fosfato.

Alcuni insegnamenti dall'adenovirus

Il gruppo di Perry, quello di Shatkin e il mio (oggi alla Rockefeller University) hanno cercato di determinare se la testa viene aggiunta all'interno del nucleo, come avevamo fatto quando si era trovato che la coda poli-*A* era una componente comune degli m-RNA cellulari. Abbiamo così scoperto che le teste sono aggiunte all'estremità libera 5' del filamento di RNA che si sta allungando a opera di enzimi presenti nel nucleo prima che la

La dimostrazione di un processo di maturazione si è avuta per la prima volta con l'r-RNA. Nella cellula di mammifero la maggior parte dell'RNA stabile si trova nel citoplasma; una quantità molto più piccola di esso è presente, invece, nel nucleolo, un organello del nucleo. La dimensione di una molecola di RNA si misura in base alla velocità con cui precipita in una provetta quando viene centrifugata ad alta velocità; l'unità di grandezza determinata in questo modo è lo Svedberg (*S*). I tre tipi di RNA stabile sono gli r-RNA 18*S* e 28*S* e i più piccoli t-RNA. Negli esperimenti, i precursori nucleotidici radioattivi possono venir incorporati in nuovo RNA e le molecole marcate possono dare informazioni sulla dimensione dell'RNA trascritto. Negli esperimenti sulla maturazione dell'RNA, cellule di mammifero sono state messe in presenza di precursori radioattivi (*1*). Dopo cinque minuti questi marcatori sono comparsi in RNA di varia lunghezza nel nucleo (*2*). Dopo 25-30 minuti, buona parte del marcatore è stata trovata in una molecola nucleare delle dimensioni di 45*S* (*3*). L'antibiotico actinomicina, che arresta la sintesi dell'RNA, è stato somministrato alle cellule marcate (*4*): a questo punto il marcatore è scomparso dalle molecole 45*S* ed è comparso, invece, nel citoplasma nelle molecole 18*S* e 28*S*. Pertanto l'RNA 45*S* è risultato un precursore che viene accorciato per dare, prima del trasporto nel citoplasma, i tipi stabili di r-RNA.

polimerasi abbia trascritto una ventina circa di basi. Il gruppo di Moss, invece, ha isolato di recente l'enzima che aggiunge la testa. A differenza della coda poli-*A*, questa viene aggiunta a tutti gli m-RNA trascritti. La funzione della testa non è completamente chiara, ma sembra che essa serva a promuovere la traduzione. Dato che viene aggiunta a tutti gli m-RNA, non pare che l'addizione possa essere una forma significativa di regolazione genetica.

La scoperta della poliadenilazione e dell'aggiunta della testa ha indirizzato di nuovo la nostra attenzione alla relazione esistente tra l'hn-RNA e l'm-RNA. Si sapeva già che i due tipi di molecola hanno una composizione in basi simile. Quando l'hn-RNA e l'm-RNA poliribosomale vennero purificati mediante cromatografia per affinità, si trovò che in

Il precursore dell'RNA ribosomale nelle cellule eucarioti include tre «spaziatori» che vengono rimossi nel processo di maturazione. Metodi avanzati per determinare le sequenze nucleotidiche del DNA e dell'RNA hanno mostrato come l'RNA primario, che nelle cellule umane ha un valore di sedimentazione di 45*S*, viene trasformato negli r-RNA 18*S* e 28*S*. La struttura generale di quasi tutti gli RNA ribosomali di cellule eucarioti, prodotti primari di trascrizione, è simile: quello illustrato qui proviene da cellule tumorali umane in coltura, dette cellule HeLa. Nel primo stadio della maturazione intervengono gli enzimi che provocano una scissione tra i segmenti 18*S* e 5,8*S*. Gli spaziatori sono poi rimossi da altri enzimi; l'RNA 5,8*S* si attacca all'RNA 28*S* mediante legami a idrogeno, producendo le due forme stabili di RNA presenti nei ribosomi.

Una «coda» e una «testa» sono aggiunte all'RNA trascritto dal DNA nelle prime fasi di maturazione dell'm-RNA. Quando la RNA-polimerasi ha trascritto solo circa 20 nucleotidi, gli enzimi presenti nel nucleo aggiungono una struttura molecolare all'estremità 5' libera del prodotto di trascrizione (*1, 2*). Questa testa consiste di una guanina (*G*) modificata, attaccata al prodotto suddetto mediante un legame che include tre gruppi fosfato (*P*). Allorché la polimerasi si sposta più lontano lungo il DNA, trascrive la sequenza *AAUAAA*, che svolge un ruolo nell'aggiunta della serie di nucleotidi *A*, chiamata poli-*A* (*3*). Il prodotto di trascrizione viene tagliato circa 20 basi dopo la sequenza *AAUAAA* (*4*). Lo stesso enzima, o un altro, aggiunge quindi i 150-200 nucleotidi *A* del poli-*A* all'estremità 3' della catena (*5*). In fasi successive del processo di maturazione, le strutture terminali si conservano, mentre la parte interna dell'RNA trascritto si modifica.

ambedue i gruppi le molecole avevano una testa 5' e una coda poli-*A* 3'. Pertanto se l'hn-RNA è il precursore dell'm-RNA, le estremità della molecola vengono preservate nel corso della maturazione e del trasporto nel citoplasma. Però, gli hn-RNA erano in media cinque volte più lunghi degli m-RNA poliribosomali. Una possibile spiegazione di questa differenza sembra intuitivamente ovvia: un segmento potrebbe essere tagliato dal centro dell'hn-RNA e ciò che resta della molecola potrebbe poi venir saldato di nuovo lasciando intatte le estremità. Tuttavia, nel 1976, l'idea sembrava di difficile realizzazione o addirittura irrealizzabile agli esperti di biologia molecolare.

In meno di due anni, invece, risultò che tale lontana possibilità aveva effettivamente luogo in una cellula eucariote. Buona parte delle ricerche effettuate per dimostrare che l'hn-RNA viene accorciato per diventare m-RNA ha utilizzato come parte del sistema sperimentale l'adenovirus, uno dei molti virus che può provocare infezioni delle vie respiratorie superiori nell'uomo. Il decorso dell'infezione da adenovirus è stato studiato a fondo. In particolare sono oggi ben comprese la collocazione cronologica della duplicazione del DNA virale nel nucleo della cellula, la composizione delle proteine virali e la sequenza secondo la quale le proteine sono prodotte. Pertanto l'adenovirus risultò un buon organismo da studiare per trovare in che modo l'RNA specifico del virus viene sintetizzato nel nucleo cellulare e fatto maturare per ottenere m-RNA.

Il genoma dell'adenovirus è una molecola di DNA a duplice filamento, di lunghezza pari a 36 000 nucleotidi. La posizione di una particolare sequenza di nucleotidi nel genoma virale è generalmente descritta in base a una mappa lineare di 100 unità in cui ogni unità corrisponde a 360 nucleotidi e dove 0 si riferisce all'estremità sinistra della mappa e 100 alla destra. Da otto a dieci ore dopo che la particella virale è penetrata in una cellula umana, la cellula comincia a produrre copie del DNA virale. Le proteine virali vengono classificate in rapporto al tempo della duplicazione del DNA: quelle «precoci» sono prodotte prima che il DNA si duplichi; il piccolo gruppo di quelle «intermedie» all'incirca nello stesso periodo in cui si duplica in DNA; infine, le «tardive» principalmente dopo che la duplicazione si è messa in corso. La maggior parte delle proteine che formano la struttura della particella virale sono proteine tardive; tra di esse vi sono le proteine degli esoni, le più abbondanti nell'involucro esterno (o capside) del virus.

Le ricerche sul processo di maturazione dell'RNA messaggero si sono concentrate sulle proteine tardive dell'adenovirus. Il lavoro sperimentale si è basato molto intensamente sulle endonucleasi di restrizione, un gruppo di enzimi che ha avuto una profonda influenza sulla biologia molecolare quando, negli anni settanta, è stato scoperto. Ogni endonucleasi di restrizione taglia il DNA o l'RNA in corrispondenza di una specifica sequenza di quattro-sei nucleotidi. In un particolare genoma, i siti di restrizione, cioè i punti in cui l'enzima taglia il DNA, sono fissi e, pertanto, le endonucleasi di restrizione possono essere utilizzate per tagliare una molecola di DNA in segmenti riproducibili. Quando i siti di restrizione sono stati identificati, i segmenti si possono raggruppare in una serie ordinata, con ogni segmento in una posizione nota sulla mappa lineare del genoma virale.

Un passo avanti nell'ibridazione

Si può vedere facilmente che il fatto di possedere una simile serie di pezzi di DNA ordinati aumenta notevolmente il potere della tecnica di ibridazione molecolare. Se le molecole marcate di RNA vengono mescolate con una serie ordinata di segmenti di DNA, gli ibridi che ne risultano indicano con grande precisione la posizione nel genoma da cui l'RNA è stato trascritto. Con questa tecnica formidabile abbiamo cercato di determinare se gli RNA messaggeri per le proteine tardive dell'adenovirus vengono trascritti nel nucleo cellulare come una singola unità lunga e quindi tagliati in segmenti più piccoli o se, invece, sono trascritti in unità che hanno all'incirca le dimensioni delle molecole del messaggero.

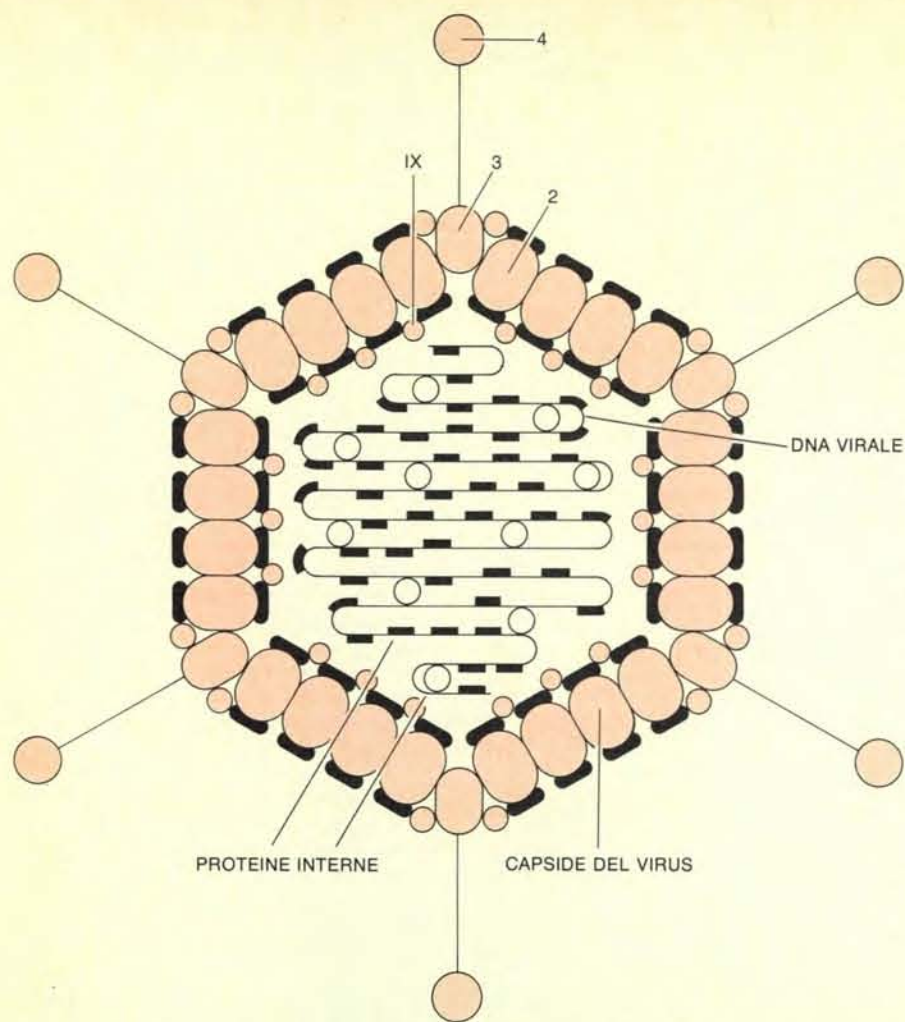
Quando la trascrizione dell'RNA virale aveva preso inizio, le cellule infettate con l'adenovirus sono state esposte a nucleosidi radioattivi per un tempo così breve che solo alcuni precursori marcati si sono aggiunti a ogni filamento di RNA. In una popolazione di cellule infettate, vi sono RNA-polimerasi pressoché in ogni posizione del DNA virale che funge da stampo per la trascrizione dell'RNA. Se alle cellule viene fornito il marcatore, alle estremità di un insieme di molecole di RNA di differenti lunghezze si troveranno nucleotidi radioattivi. Gli RNA marcati più corti si troveranno vicini al punto di inizio della trascrizione. Si osserveranno anche molecole più lunghe e la più lunga in assoluto si troverà vicino al punto terminale della trascrizione.

L'RNA marcato è stato estratto dalle cellule e purificato. I vari RNA purificati poi sono stati separati in base alle dimensioni e ibridati con una serie ordinata di segmenti di DNA di adenovirus. L'ibridazione molecolare ci ha fornito un modo per saggiare le ipotesi su come viene prodotto l'RNA messaggero. Se gli m-RNA per le proteine tardive venissero trascritti in piccole unità corrispondenti ai singoli m-RNA, l'ibridazione darebbe un quadro complesso in cui RNA corti sarebbero ibridati con il DNA vicino a ciascuno dei molteplici siti di inizio della trascrizione.

Si consideri ora che cosa accadrebbe se gli RNA messaggeri per le proteine tardive venissero trascritti come una singola unità e quindi tagliati in segmenti. Vi sarebbe soltanto un punto di inizio per la trascrizione e gli RNA più corti si ibriderebbero solo con il DNA vicino all'unico punto di inizio. Gli RNA marcati più lunghi si ibriderebbero, nell'unità di trascrizione, in corrispondenza delle giuste posizioni intermedie. Quando l'RNA nucleare brevemente marcato e proveniente dalle cellule infettate è stato ibridato con il DNA dell'adenovirus, gli RNA più corti si sono ibridati solo in corrispondenza di un'unica posizione: l'unità di mappa 16. Gli RNA più lunghi si sono dispersi, invece, in base alle loro dimensioni, alla destra dell'unità 16. Pertanto abbiamo concluso che gli RNA messaggeri per le proteine tardive vengono trascritti come unità singola a partire da una posizione vicina a quella di mappa 16.

Posizioni sulla mappa genica

Due gruppi di giovani ricercatori, uno che operava presso il Cold Spring Harbor Laboratory e l'altro al Massachusetts Institute of Technology, hanno compiuto una serie di esperimenti che hanno rivelato in che modo avviene il processo di maturazione dell'RNA primario, cioè appena trascritto dall'unità di trascrizione per la proteina tardiva. Richard J. Roberts e Richard E. Gelinas del Cold Spring Harbor Laboratory hanno trovato che tutti gli RNA messaggeri per le proteine tardive dell'adenovirus avevano gli stessi 11 nucleotidi confinanti con la testa in corrispondenza dell'estremità 5'. Philip A. Sharp, Susan M. Berget e Claire Moore del MIT, lavorando con l'm-RNA per una delle proteine degli esoni, hanno trovato che la sequenza comune adiacente alla testa non si ibridava con la porzione principale dell'm-RNA. Quando l'm-RNA per l'esone veniva mescolato con il DNA, formava in modo costante un ibrido, con lunghezza corrispondente a circa 4500 basi, nel mezzo del genoma virale che poteva essere reso visibile al microscopio elettronico. Dalla struttura centrale a duplice filamento si estendevano, tuttavia, due code. All'estremià 3' vi era la sequenza poli-*A*, dalla quale non ci si attendeva che si ibridasse perché non derivante dal DNA virale. All'altra estremità dell'm-RNA, la sequenza comune 5' si estendeva anch'essa dall'ibrido come coda a filamento singolo.

Ambedue i gruppi hanno riconosciuto rapidamente la possibilità che la sequenza 5' provenisse da un'altra parte del genoma dell'adenovirus. Sharp al MIT e, inoltre, Thomas R. Broker e Louise T. Chow al Cold Spring Harbor Laboratory hanno mescolato molti RNA messaggeri per le proteine tardive con la serie ordinata di segmenti del DNA e hanno ottenuto un risultato sorprendente: ogni m-RNA si ibridava, nel genoma virale, in quattro posizioni distinte. Ognuno formava un ibrido in un unico sito corrispondente alla sequenza codificante per la proteina. Inoltre, tutti gli m-RNA, indipendentemente dalla posizione della sequenza codificante, hanno formato degli ibridi in corrispondenza delle posizioni 16, 20 e 27. Il risultato ha suggerito che ogni m-RNA per le proteine tardive dell'adenovirus è un mosaico che consta di tre sequenze comuni con un unico brano di informazione genetica.

Un adenovirus, virus che provoca negli esseri umani infezioni delle vie respiratorie superiori, è illustrato in questa sezione trasversale schematica. Nelle prime indagini sulla maturazione dell'm-RNA, esso è stato un importante oggetto di studio, perché a quell'epoca dalle cellule di mammifero non era ancora stato isolato alcun m-RNA. L'involucro esterno dell'adenovirus consta di molecole proteiche, la maggior parte delle quali sono chiamate esoni. Le principali proteine degli esoni sono designate come 2, 3, 4 e IX. All'interno dell'involucro vi è una molecola di DNA a duplice filamento, con una lunghezza di circa 36 000 nucleotidi. Quando il virus penetra in una cellula umana, il suo DNA si duplica e da esso viene poi trascritto un numero relativamente piccolo di m-RNA. Il virus usurpa quindi l'apparato di sintesi proteica della cellula ospite per produrre proteine corrispondenti a questi m-RNA. Le proteine degli esoni fanno parte di un gruppo «tardivo» che viene sintetizzato principalmente dopo che il DNA virale ha cominciato a duplicarsi.

Ricerche recenti da parte di vari esperti, tra cui Heschel J. Raskas della School of Medicine della Washington University e Joseph R. Nevins ed Edward Ziff della Rockefeller University hanno illustrato in che modo gli m-RNA dell'adenovirus riescono a incorporare sia le sequenze comuni sia quelle uniche. In questo lavoro l'unità di trascrizione per le proteine tardive dell'adenovirus è apparsa come una pezzo estremamente complicato della macchina biologica. Si sa oggi che in una cellula di mammifero, infettata da un adenovirus, il principale prodotto di trascrizione ottenuto dopo la duplicazione del DNA è l'unità di trascrizione per le proteine tardive, che porta sequenze codificanti per le 13 o 14 proteine tardive. L'unità di trascrizione si estende dalla posizione di mappa 16 pressoché alla posizione di mappa 100, all'estremità destra della mappa stessa. Vicino all'unità 16, vi è un filamento di nucleotidi chiamato il promotore tardivo, che induce la polimerasi ad avviare la trascrizione. In corrispondenza delle posizioni 16, 20 e 27 vi sono le sequenze comuni, osservate da Sharp, Broker e Chow. Dalla posizione 27 alla posizione 100 vi sono le sequenze codificanti.

Le sequenze codificanti nell'unità di trascrizione tardiva sono disposte in cinque gruppi, designati come *L*1, *L*2, *L*3, *L*4 e *L*5 (nell'ordine da sinistra a destra sulla mappa). Ogni gruppo numerato include le sequenze codificanti per una famiglia di proteine affini. All'estremità destra di ogni gruppo vi è un sito di attacco poli-*A*. Per esempio il gruppo *L*2 si estende da circa la posizione 40 alla posizione 50; il suo sito poli-*A* è in corrispondenza della posizione 50. Così vi sono, in corrispondenza della posizione 16, un sito che funge da testa e, nell'unità di trascrizione tardiva, un totale di cinque poli-*A*.

Lo «splicing» del messaggero

Per ogni famiglia di proteine, le sequenze codificanti hanno una disposizione analoga, che si sovrappone. Tutte le sequenze codificanti si sovrappongono all'estremità destra dell'unità numerata, e tutte includono i nucleotidi adiacenti al sito poli-*A*. Nel mezzo del gruppo, invece, le sequenze codificanti divergono: ciascuna include un differente segmento dei nucleotidi che si trovano nel mezzo dell'unità. Quando una molecola matura di RNA messaggero si forma dall'unità di trascrizione tardiva viene inclusa una sola sequenza codificante.

Da molti esperimenti è risultato chiaro che l'm-RNA maturo viene prodotto nel seguente modo. La trascrizione è avviata in corrispondenza del promotore tardivo e si muove verso destra a partire dalla posizione 16. Quando alcuni nucleotidi sono stati trascritti, l'estremità 5' ondeggiante dell'RNA viene completata. La polimerasi quindi si sposta verso la posizione 100, superando tutti e cinque i siti poli-*A*. Durante il loro passaggio gli enzimi tagliano il prodotto di trascrizione in corrispondenza di un sito poli-*A* e attaccano qui la coda poli-*A*. La prima molecola di RNA negli stadi della maturazione include, pertanto, tutte le sequenze dalla posizione 16 al sito poli-*A* prescelto. Per esempio, se viene prescelto il sito *L*2, l'RNA includerà le sequenze dalla posizione 16 alla posizione 50; la famiglia *L*1 di sequenze codificanti viene inclusa con la famiglia *L*2, ma non con le famiglie *L*3, *L*4 e *L*5. L'RNA a valle del sito *L*2 (che include i gruppi *L*3, *L*4 e *L*5) viene chiaramente distrutto nel nucleo e non viene mai utilizzato come messaggero.

L'RNA nucleare, contenente il poli-*A* e parzialmente maturato, include pertanto parecchi «spaziatori», che non fanno parte dell'm-RNA maturo. Gli spaziatori vengono rimossi da enzimi in un processo che può richiedere anche 20-30 minuti. Vengono asportate anche le sequenze *L*1 e vengono rimossi i segmenti compresi tra la posizione 16 e la posizione 20 e tra la posizione 20 e la posizione 27; rimangono così solo le sequenze comuni in corrispondenza delle posizioni 16, 20 e 27. Ne risulta una molecola con un gruppo di sequenze codificanti affini. In questo caso è stata selezionata la sequenza *L*2. La famiglia della *L*2 ha tre sequenze codificanti. Una delle tre viene scelta da enzimi che rimuovono un segmento variabile di RNA dall'interno della famiglia di *L*2. Si ottiene così m-RNA maturo, con la testa costituita dalle sequenze 5' comuni, la coda poli-*A* e una sequenza codificante solamente.

I nucleotidi dalla posizione 16 alla posizione 100 nel genoma dell'adenovirus costituiscono un esempio estremo di unità di trascrizione complessa. Questa unità può subire un processo di trasformazione per dare alla fine più di un tipo di messaggero maturo. D'altra parte, un'unità di trascrizione semplice fornisce solamente un tipo di m-RNA. Grazie alla combinazione di molteplici siti poli-*A* e di uno *splicing* (rottura e successiva saldatura) differenziale, si può trasformare, partendo da una singola unità di trascrizione dell'adenovirus, una formidabile serie di proteine tardive.

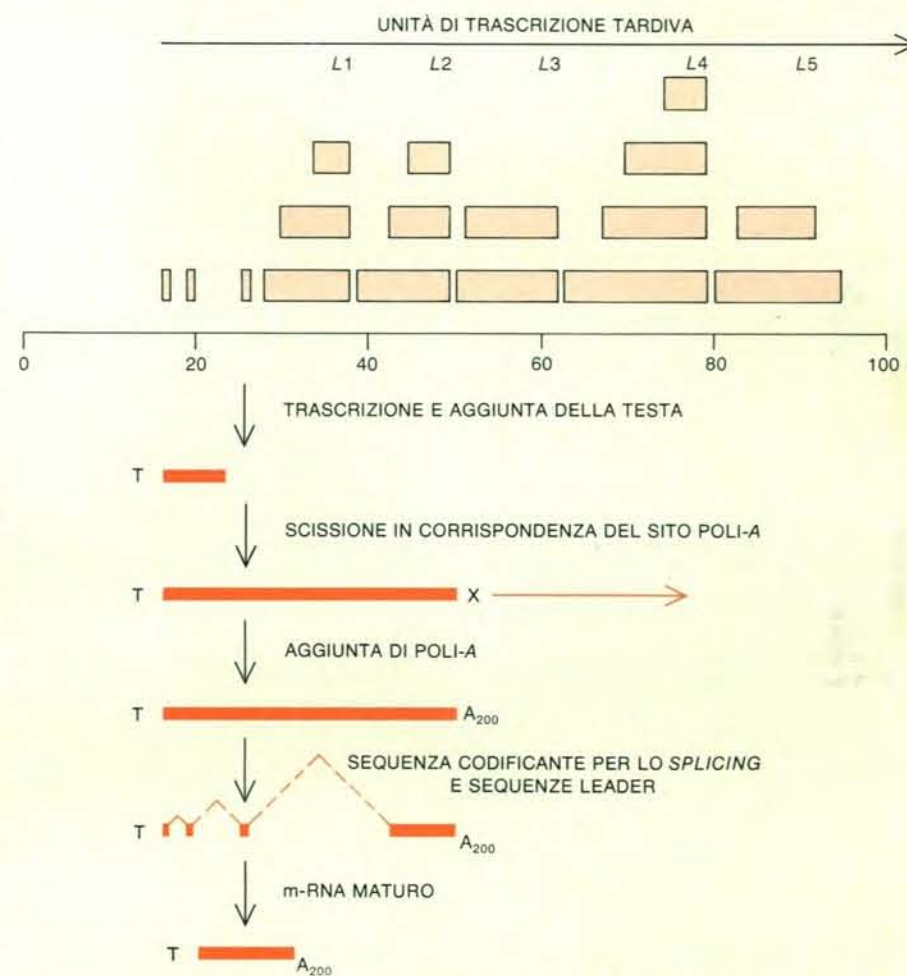
Si può vedere facilmente che o i molteplici siti poli-*A* o lo *splicing* differenziale da soli potrebbero portare a una unità di trascrizione complessa. Solo poche di queste unità complesse sono state trovate nelle cellule eucarioti e di esse tutte tranne una hanno soltanto siti poli-*A* molteplici. Soltanto una ha sia i siti poli-*A* molteplici sia lo *splicing* differenziale: essa è, nel moscerino della frutta (*Drosophila*), l'unità per la catena pesante della proteina miosina.

Nel genoma del ratto, una singola unità di trascrizione con due siti poli-*A* è stata trovata da Ronald Evans, Michael Rosenfeld e collaboratori al Salk Institute for Biological Studies. Essa include la sequenza codificante per due proteine: l'ormone calcitonina, sintetizzato nella tiroide, e un neuropeptide recentemente scoperto, un ormone del sistema nervoso sintetizzato nell'ipofisi. Le due proteine hanno teste (l'estremità della catena proteica definita ammino-terminale o N-terminale) identiche, ma code (l'estremità definita carbossiterminale o C-terminale) diverse. Nelle cellule della tiroide viene selezionato il primo sito poli-*A* e si ottiene la calcitonina; nelle cellule dell'ipofisi viene selezionato il secondo sito poli-*A* e viene prodotto il neuropeptide.

Le ricerche effettuate dal 1975 a oggi hanno dimostrato che i geni nelle cellule eucarioti possono venir regolati attraverso il processo di maturazione dell'RNA. Non è ancora chiaro, tuttavia, quanto significativo risulterà essere tale processo come forma di regolazione genetica. Per la maggior parte dei geni degli eucarioti il metodo di regolazione principale risulta essere ancora il controllo della trascrizione. Inoltre, la maggioranza delle unità di trascrizione eucarioti che sono state esaminate sono semplici unità, che possono codificare solo per una proteina.

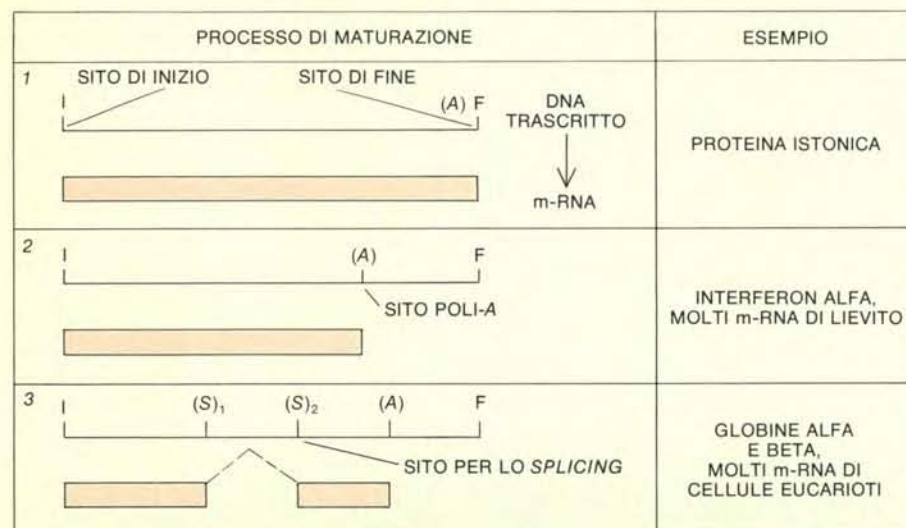
Esoni e introni

Anche in semplici unità di trascrizione, però, si svolge un processo di lavorazione che serve a eliminare gli spaziatori. Dopo che la ricerca con l'adenovirus ha tracciato il primo solco nel terreno, è risultato chiaro che, nei geni degli eucarioti, le sequenze espresse (esoni) sono anch'esse comunemente interrotte da sequenze nucleotidiche che non vengono tradotte (introni). Geni discontinui sono stati trovati in eucarioti così poco affini tra loro come i lieviti e gli esseri umani. In tutte le cellule è necessario che si verifichi uno *splicing* perché si formi un RNA messaggero senza gli introni e con gli esoni uniti per la traduzione. Così la maturazione dell'RNA è molto diffusa tra gli eucarioti. I geni discontinui degli eucarioti rappresentano una notevole innovazione rispetto ai geni dei procarioti. In una cellula procariote, i geni sono continui e segmenti supplementari di DNA vengono subito scartati dal genoma. Le opinioni un tempo prevalenti sull'evoluzione cellulare sostenevano che gli eucarioti si erano evoluti da organismi unicellulari molto simili alle attuali cellule procarioti. Se ciò è giusto, ne consegue che gli introni vennero introdotti nei geni dei procarioti nel corso dell'evoluzione. D'altra parte, Carl R. Woese e collaboratori dell'Università dell'Illinois a Urbana-Champaign hanno avanzato l'ipotesi che gli eucarioti non siano derivati dai procarioti e che vi siano due linee di discendenza procarioti ugualmente antiche. Woese sostiene anche che le tre linee di discendenza siano derivate da un unico precursore, che egli chiama «progenote».

L'analisi particolareggiata degli introni nei geni degli eucarioti conferma l'idea che gli eucarioti non si siano evoluti dai procarioti. Innanzitutto le sequenze intercalate in alcuni geni degli eucarioti esistono al loro posto da moltissimo tempo. Per esempio, nell'emoglobina, una molecola che trasporta ossigeno, vi sono due catene proteiche: la globina alfa e la globina beta. Le sequenza amminoacidiche delle due catene sono diverse, ma le due catene sono comunque ripiegate in modo da formare, nella molecola dell'emoglo-

L'unità di trascrizione tardiva dell'adenovirus ha sia siti poli-*A* multipli sia siti multipli per lo *splicing*. La si può trasformare così che dia degli m-RNA che codificano per più di una dozzina di proteine tardive dell'adenovirus. Una unità di trascrizione è un filamento di DNA che viene continuamente trascritto. L'unità tardiva dell'adenovirus si estende dalla posizione 16 praticamente fino alla posizione 100 su una mappa lineare del DNA virale, in cui ogni unità corrisponde a 360 nucleotidi. In tutti gli m-RNA tardivi vi sono sequenze complementari alle posizioni 16, 20 e 27. Dalla posizione 30 fino a quasi la posizione 100 vi sono sequenze che codificano per le proteine tardive. raggruppate in famiglie designate con *L*1, *L*2, *L*3, *L*4 e *L*5. Ogni famiglia include le regioni codificanti per parecchie proteine affini. All'estremità destra di ogni famiglia vi è un singolo sito poli-*A*, comune a tutti gli m-RNA presenti in quel gruppo. Subito dopo che ha inizio la trascrizione, all'estremità 5' si aggiunge una struttura, detta testa (*T*); per finire la RNA-polimerasi trascrive l'intera unità tardiva. Allorché attraversa i cinque siti poli-*A*, uno di essi viene prescelto; altri enzimi tagliano via in questo punto i nucleotidi e viene aggiunto il poli-*A*. Qui è stata presa in esame l'*L*2: le sequenze tra le posizioni 16, 20 e 27 e tra la posizione 27 e la famiglia *L*2 vengono rimosse. Uno dei segmenti codificanti della famiglia *L*2 viene quindi prescelto; non è nota però la base della scelta per il sito poli-*A* e per il sito in cui avviene lo *splicing*. L'm-RNA maturo include la testa aggiunta in corrispondenza dell'estremità 5', le sequenze comuni nelle posizioni 16, 20 e 27, il codice per una particolare proteina tardiva e la coda all'estremità 3'.

bina, analoghe strutture tridimensionali. Tali strutture sono anche simili alla struttura della mioglobina, la proteina che si lega all'ossigeno e che è presente nelle cellule muscolari.

Malgrado queste somiglianze strutturali, le differenze nelle sequenze amminoacidiche della globina alfa, della globina beta e della mioglobina fanno pensare che i geni per le tre proteine abbiano cominciato a divergere da un precursore comune forse un miliardo di anni fa. Nel contempo, tutti e tre i geni hanno tre esoni separati da due introni che si trovano in posizioni analoghe. Laddove le sequenze degli amminoacidi hanno cominciato a divergere, sembra che la disposizione degli introni si sia conservata, cosicché essi hanno potuto rimanere al loro posto per un miliardo di anni.

PROCESSO DI MATURAZIONE	ESEMPIO
1 SITO DI INIZIO I — SITO DI FINE (A) F; DNA TRASCRITTO → m-RNA	PROTEINA ISTONICA
2 I (A) F; SITO POLI-A	INTERFERON ALFA, MOLTI m-RNA DI LIEVITO
3 I (S)$_1$ (S)$_2$ (A) F; SITO PER LO *SPLICING*	GLOBINE ALFA E BETA, MOLTI m-RNA DI CELLULE EUCARIOTI

Una semplice unità di trascrizione può essere trasformata in modo da dare un solo tipo di m-RNA. A sinistra, viene rappresentato in forma schematica il processo di maturazione di unità semplici, mentre a destra sono riportati esempi di tali unità e organismi nei quali si trovano. Il sito di inizio è dove ha inizio la trascrizione; il sito di fine è dove termina la trascrizione. Il sito poli-*A* corrisponde al punto in cui la coda 3' di poli-*A* viene inserita. In alcuni casi, la RNA-polimerasi si arresta in corrispondenza del sito poli-*A* (*1*). In altri casi, invece, la trascrizione procede al di là di esso (*2*). Il prodotto di trascrizione può venir tagliato e saldato così da dare un m-RNA finito di minore lunghezza, se nel suo interno si trovano delle sequenze non codificanti (*3*). La grande maggioranza delle unità di trascrizione trovate nelle cellule dei mammiferi sono semplici.

PROCESSO DI MATURAZIONE	ESEMPIO
1 SITO DI INIZIO I (A)$_1$ (A)$_2$ SITO DI FINE F	CATENE PESANTI DELL'IMMUNOGLOBULINA, CALCITONINA
2 I (S)$_1$ (S)$_2$ (S)$_3$ (A) F	GENI IN PREVALENZA VIRALI: SV 40, POLIOMA, RETROVIRUS

Un'unità di trascrizione complessa può venir trasformata così da dare più di un m-RNA e pertanto più di una sola proteina. La presenza di due o più siti poli-*A* può sfociare in una unità complessa (*1*). Se questa viene tagliata in corrispondenza del primo sito poli-*A*, si produce un solo m-RNA; se viene tagliata, invece, in corrispondenza di un altro sito, si produce un differente m-RNA. In una unità di trascrizione con un solo sito poli-*A*, vi possono essere più di un unico sito per lo *splicing*. L'm-RNA che viene prodotto dipende da quali siti per lo *splicing* siano prescelti (*2*). Unità complesse con più di un sito poli-*A* sono comuni nei virus e sono state trovate anche nelle cellule di mammifero. Unità con *splicing* differenziale sono state notate in virus e un esempio anche in una cellula eucariote. Unità di trascrizione complesse, con molteplici siti per lo *splicing*, possono venir integrate nei geni della cellula ospite ed espresse assieme ai geni di quest'ultima.

Inoltre, quando si esamina la struttura tridimensionale di una proteina, si trova spesso che la catena proteica può essere suddivisa in segmenti con differenti funzioni; tali segmenti prendono il nome di campi funzionali. Molti introni dividono i geni in vari pezzi, ciascuno dei quali codifica per un differente campo funzionale. Pertanto, se gli introni venissero introdotti a caso in un genoma preesistente, avrebbero il notevole effetto di dividere i geni in pezzi corrispondenti a funzioni separate. Sembra più probabile che essi si trovassero già al loro posto forse fin dall'inizio dell'evoluzione e che gli esoni codificanti per i campi funzionali si siano evoluti in maniera indipendente l'uno dall'altro.

All'inizio era l'RNA

Sulla base di queste argomentazioni e di altre ancora, W. Ford Doolittle della Dalhousie University nella Nuova Scozia, Darryl C. Reanney della Latrobe University in Australia e io stesso abbiamo proposto che, anche in tempi precellulari, l'informazione per le proteine non sia stata continua. Si ritiene da più parti che, anche se il DNA viene prima dell'RNA nella sequenza che porta alla sintesi delle proteine nelle cellule attuali, il primo acido nucleico codificante non fu DNA, ma RNA. Tra le varie ragioni si annovera il fatto che le primitive catene di RNA possono essere sintetizzate senza enzimi. Inoltre, l'RNA può immagazzinare informazioni codificanti ed è necessario per la traduzione. Se l'RNA si è evoluto come catena nucleotidica casuale, la sua informazione utile potrebbe non essere stata continua. Lo *splicing* dell'RNA per riunire pezzi d'informazione discontinui sarebbe stato così un evento evolutivo precoce e propizio. Di recente, Thomas R. Cech dell'Università del Colorado a Boulder ha dimostrato che l'RNA può essere tagliato e saldato di nuovo senza il concorso di enzimi, perlomeno nel protozoo *Tetrahymena*. Pertanto non è da escludere che le sequenze intercalari esistessero fin dai tempi più antichi e che lo *splicing* dell'RNA sia stato una caratteristica precoce dell'espressione genica.

Queste congetture sull'evoluzione non troveranno subito una soluzione. Nel frattempo rimane da rispondere ad alcuni pressanti e precisi interrogativi. Si sta compiendo uno sforzo molto intenso per identificare gli enzimi che tagliano e risaldano il prodotto di trascrizione. Quando si troveranno, sarà possibile scoprire come un tipo di *splicing* venga prescelto al posto di un altro. Analogamente si dovrà anche conoscere perché un sito poli-*A* venga selezionato al posto di un altro. A un livello più generale si pone la questione di sapere quanto sia significativo il processo di maturazione dell'RNA nella regolazione genetica in complesso. Di qualunque natura saranno le risposte, le indagini sulla maturazione dell'RNA continueranno ad avere un effetto profondo sul nostro modo di capire in che modo i geni si esprimano nelle cellule eucarioti.

La neuropsicologia della memoria a breve termine

Lo studio di pazienti con lesioni cerebrali localizzate conferma che i due sistemi di memoria, a breve e a lungo termine, oltre a essere funzionalmente distinti, sono mediati da strutture nervose diverse

di Giuseppe Vallar

L'idea che la memoria umana possa essere distinta in componenti funzionali discrete risale almeno al XIX secolo. Ad esempio, nel 1890 l'americano William James, nel suo libro *The principles of psychology,* usava i termini «memoria primaria» e «memoria secondaria» per indicare due componenti separate della memoria. Egli diede dei due tipi di memoria una definizione introspettiva: un evento nella memoria primaria non ha mai lasciato la coscienza ed è parte del presente psicologico, mentre un evento richiamato dalla memoria secondaria è stato assente dalla coscienza e fa parte del passato psicologico. Tuttavia, questa concezione a due fasi dei processi di memoria, pur essendo intuitivamente molto attraente, fu presa in ben scarsa considerazione negli anni della psicologia behaviorista.

Solo dalla metà degli anni cinquanta ricerche psicologiche sempre più numerose hanno fornito un supporto sperimentale all'intuizione introspettiva di James che la memoria umana non è un processo unitario, ma può essere frazionata in più componenti: la memoria primaria o a breve termine e la memoria secondaria o a lungo termine. Per una corretta valutazione di questa dicotomia va tenuto presente che, nella grande maggioranza, gli esperimenti condotti sui soggetti normali hanno utilizzato materiale verbale quale lettere, parole o frasi; a essi fa riferimento il presente articolo. Gli studi condotti con materiale di tipo non verbale sono assai meno numerosi.

Con questa premessa, le proprietà funzionali delle due componenti possono essere così sintetizzate. La memoria a breve termine permette di ricordare, per periodi di tempo brevi (nell'ordine dei secondi), quantità limitate d'informazione. La memoria a lungo termine assicura il ricordo di quantità maggiori d'informazione per periodi di tempo molto lunghi, senza limiti ben definiti (anni o addirittura tutta la vita). Ripetere un numero telefonico dopo che ci è stato detto è un classico compito che impegna la memoria a breve termine verbale. Ricordare ciò che abbiamo fatto ieri, o dieci anni fa, imparare una poesia a memoria, studiare il contenuto di un testo professionale sono invece situazioni in cui è impegnata la memoria a lungo termine.

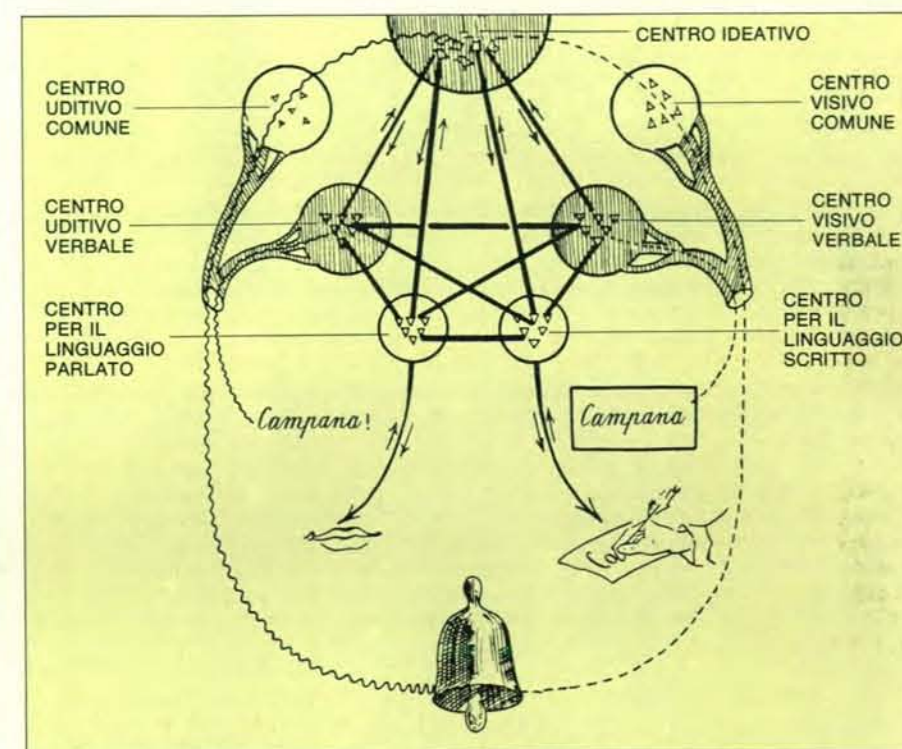
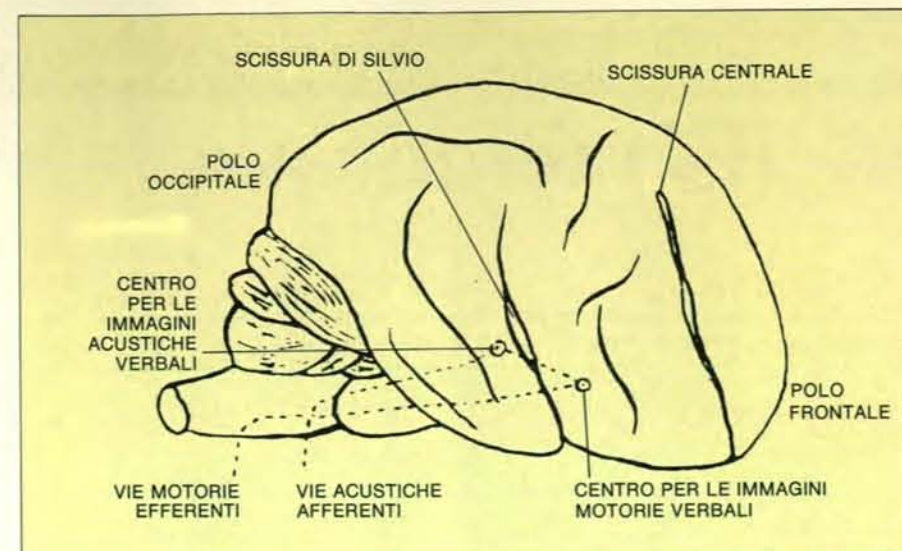
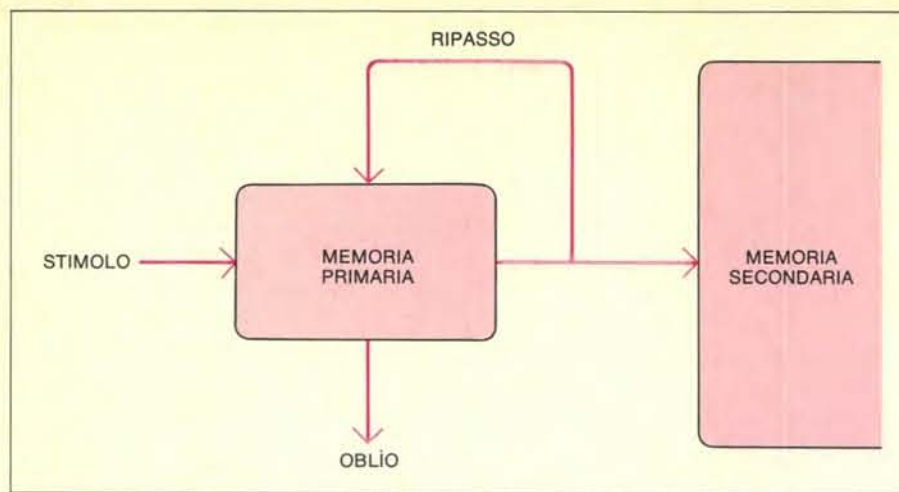
Secondo i modelli psicologici elaborati alla fine degli anni sessanta, la memoria a breve termine può essere immaginata come un «magazzino» che può contenere una quantità limitata d'informazione per un limitato periodo di tempo. Il suo contenuto può avere due destini: essere dimenticato nel giro di pochi secondi oppure esser avviato alla più durevole memoria a lungo termine. È anche possibile mantenere vivida l'informazione nella memoria a breve termine mediante una forma di «ripasso», verosimilmente equivalente all'operazione di ripetere tra sé e sé l'informazione che si vuole ricordare.

La memoria a lungo termine, invece, può essere descritta come un magazzino a capacità molto ampia, i cui limiti non sono a tutt'oggi ben noti. L'oblio dell'informazione contenuta in questa forma di memoria è molto più lento di quello dell'informazione contenuta nella memoria a breve termine. Le due memorie differiscono anche per le modalità secondo le quali l'informazione è in esse rappresentata. Nella memoria a breve termine è prevalente una codificazione che tende a preservare le caratteristiche fisiche dello stimolo: ad esempio, informazioni uditive o visive sarebbero rappresentate rispettivamente secondo un codice acustico o visivo. Nella memoria a lungo termine, invece, è dominante una rappresentazione di tipo semantico: la codificazione dell'informazione tiene in scarsa considerazione le caratteristiche fisiche dello stimolo, privilegiando invece il suo significato. Ad esempio, le parole «manto», «santo» e «vanto» sono confondibili nella memoria a breve termine in quanto simili fonologicamente, ma non nella memoria a lungo termine, perché diverse quanto a significato. Al contrario, le parole «grande», «enorme», «immenso» possono esser più facilmente confuse nella memoria a lungo termine che in quella a breve termine: infatti si tratta di sinonimi dissimili da un punto di vista fonologico.

I risultati di molti esperimenti condotti su soggetti normali sono compatibili con la dicotomia tra memoria a breve termine e memoria a lungo termine. Tuttavia, le equazioni «memoria a breve termine = rappresentazione dell'informazione secondo un codice fisico» e «memoria a lungo termine = rappresentazione dell'informazione secondo un codice semantico» sono una semplificazione eccessiva. In effetti, è stato osservato che il materiale verbale può esser immagazzinato nella memoria a lungo termine anche secondo un codice fonologico: questa possibilità è dimostrata, ad esempio, dal fatto che soggetti normali possono produrre, richiamandole dalla memoria a lungo termine, sequenze di parole che fanno rima tra loro. Per converso, fattori semantici possono esser importanti nella memoria a breve termine.

L'esigenza di rispondere a questo tipo di problemi è ben evidente nei modelli che distinguono tra livelli diversi a cui è analizzata l'informazione, ma non tra componenti distinte in cui è contenuta l'informazione. Secondo questa impostazione, suggerita nel 1972 dagli psicologi Fergus Craik e Robert Lockhart dell'Università di Toronto, in Canada, l'informazione può esser elaborata a livelli «superficiali» o «profondi» in relazione alle esigenze della situazione. Per una ritenzione temporanea dell'informazione è sufficiente un'analisi superficiale, basata sulle caratteristiche fisiche dello stimolo. Un apprendimento stabile richiede, invece, un'analisi a livello profondo, ove gli aspetti semantici sono preponderanti. Non esistono, con questo modo di vedere, magazzini distinti atti a contenere l'informazione, ma un unico processo centrale che controlla l'analisi dell'informazione secondo livelli diversi di elaborazione. Quanto più l'analisi sarà profonda, tanto più il ricordo sarà duraturo.

La neuropsicologia, la disciplina che studia i disturbi cognitivi causati da lesioni cerebrali, ha fornito qualche risposta a questi problemi. In primo luogo, lo studio di pazienti con lesioni cerebrali focali (localizzate) ha corroborato la distinzione tra i due tipi di memoria; inoltre, ha permesso di precisare meglio le caratteristiche funzionali dei due sistemi. Lo studio neuropsicologico è utile, in quanto la dimostrazione che uno o più pazienti con lesioni cerebrali focali presentano un difetto selettivo di una delle componenti del sistema costituisce una prova dell'esistenza di quella componente come entità funzionale distinta. Così, se la teoria a due componenti dei sistemi di memoria è vera, è prevedibile che esistano pazienti in cui ora l'uno ora l'altro sistema è selettivamente compromesso.

In effetti, è noto da molti anni che lesioni di particolari regioni del cervello (l'ippocampo, i corpi mammillari, alcuni nuclei del talamo) possono produrre un difetto della memoria a lungo termine, che si associa a una memoria a breve termine normale. L'esistenza di malati di questo tipo, secondo gli psicologi Richard C. Atkinson e Richard M. Shiffrin della Stanford University in California, è una prova molto importante a favore dei modelli di memoria a due stadi (si veda a questo proposito l'articolo *Il controllo della memoria a breve termine* di Richard C. Atkinson e Richard M. Shiffrin in «Le Scienze», n. 39, novembre 1971). In anni più recenti (la prima osservazione, del 1969, è dei neuropsicologi britannici Elizabeth Warrington e Tim Shallice del National Hospital for Nervous Diseases di Londra), sono stati descritti pazienti che presentavano la dissociazione opposta: un difetto della memoria a breve termine uditiva verbale, associato a una memoria a lungo termine normale. I due gruppi di soggetti differiscono non soltanto per il tipo di difetto di memoria, ma anche per la diversa sede della lesione cerebrale. Nei pazienti in cui la memoria a lungo termine è compromessa, il danno interessa alcune strutture anatomiche profonde, site sotto la corteccia cerebrale. Nei pazienti in cui la memoria a breve termine uditiva verbale è difettosa, la lesione è situata, invece, nell'emisfero cerebrale sinistro, dominante per il linguaggio nei soggetti destrimani; le regioni profonde, il cui danno produce difetti della memoria a lungo termine, sono risparmiate. L'osservazione che parti diverse del cervello producono disturbi di memoria differenti corrobora la distinzione tra memoria a breve termine e memoria a lungo termine, suggerendo una dissociazione non soltanto funzionale, ma anche strutturale: regioni anatomiche distinte del cervello sono importanti per diverse componenti dei processi di memoria.

La psicologia cognitiva fa oggi largo uso di schemi costituiti da componenti discrete, funzionalmente collegate tra loro. In tale rappresentazione del flusso dell'informazione è implicito un principio caratteristico della psicologia cognitiva, il «principio della modularità», secondo cui ogni computazione può essere suddivisa in diverse parti, relativamente indipendenti l'una dall'altra. Nell'ultimo decennio si è molto sviluppato un indirizzo di ricerca neuropsicologica, che ha utilizzato questi modelli modulari nello studio dei difetti cognitivi di pazienti con lesioni cerebrali. Presumendo che l'indipendenza computazionale sia sottesa da processi fisici cerebrali potenzialmente separabili (ad esempio, localizzati in parti diverse del cervello), si può prevedere che una lesione cerebrale danneggi selettivamente un particolare sottosistema. Ciò produrrà una dissociazione tra la prestazione in un compito, che fa uso prevalente della componente danneggiata, e quella in altri compiti, che utilizzano altri sottosistemi. Mentre la prima sarà compromessa, le seconde dovrebbero essere indenni. Nell'articolo è trattato il caso dei processi di memoria a breve e a lungo termine; tuttavia, i modelli modulari sono stati utilizzati per lo studio di altri difetti neuropsicologici, ad esempio le dislessie. L'utilizzazione in neuropsicologia di schemi di questo genere è ben più antica della psicologia cognitiva: infatti risale alla seconda metà del secolo scorso. La figura in alto mostra l'organizzazione cerebrale del linguaggio secondo uno schema del neurologo tedesco Carl Wernicke (1848-1905). Lesioni delle aree del linguaggio possono causare afasia, ma il quadro clinico varia secondo la sede della lesione. La figura in basso mostra lo schema, cosiddetto «della campana», del neurologo francese Jean-Martin Charcot (1825-1893). Tale schema mostra come, secondo la concezione di Charcot, possa avvenire il riconoscimento di un oggetto (nel suo esempio, una campana) presentato visivamente, oppure del suo suono. Lo schema illustra, inoltre, il processo di comprensione uditiva e di lettura di parole. Infine, sono descritte le possibilità di risposta parlata e scritta (nell'esempio, la denominazione della campana). Gli schemi dei neurologi dell'Ottocento, di cui quelli qui illustrati sono alcuni esempi, localizzano le diverse componenti funzionali del linguaggio in aree cerebrali distinte. La somiglianza tra questi modelli e quelli degli psicologi cognitivisti contemporanei è evidente.

Viene qui rappresentato il modello a due stadi, sviluppato nel 1965 da Nancy Waugh della Harvard Medical School e Donald Norman della Harvard University, che usarono la terminologia di William James (1890). Secondo loro, ogni stimolo verbale a cui si presta attenzione entra nella «memoria primaria»: questa ha una capacità limitata e quindi nuovi stimoli si sostituiscono ai vecchi, che vanno perduti. Tuttavia, quando un'informazione viene ripassata (per il materiale verbale questo corrisponde spesso alla ripetizione dell'informazione tra sé e sé), ritorna nella memoria primaria e può essere avviata alla più duratura «memoria secondaria», la cui capacità è molto maggiore.

Il difetto più marcato, che si manifesta in pazienti con memoria a breve termine uditiva verbale compromessa e che venne riferito fin dalla prima osservazione della Warrington e di Shallice, consiste nella incapacità a ripetere sequenze di stimoli verbali immediatamente dopo la loro presentazione. Mentre un soggetto normale riesce a ripetere senza problemi una serie costituita da cinque o più numeri, questi pazienti sono già in difficoltà nel ripetere serie di due o tre numeri, lettere o parole. Il difetto non può esser spiegato in termini di una non buona percezione dello stimolo uditivo in quanto questa è perfetta. Anche un'interpretazione che attribuisca il difetto a un qualche problema nel produrre la risposta, ad esempio a disordini di tipo articolatorio, non è accettabile: il difetto di memoria a breve termine, infatti, non si riduce se i pazienti rispondono indicando gli stimoli verbali che sono stati appena detti scegliendoli tra alternative, cioè fornendo una risposta di tipo non verbale.

Nella memoria a breve termine di questi pazienti esiste un'importante dissociazione in relazione alla modalità sensoriale in cui l'informazione verbale viene presentata. La memoria a breve termine è migliore se l'informazione viene presentata nella modalità visiva. La dissociazione uditivo-visiva permette di qualificare il difetto come specifico della modalità uditiva e suggerisce nel contempo che esistono magazzini a breve termine separati, rispettivamente per informazioni visive e uditive. L'osservazione che la memoria a breve termine per suoni significativi non verbali (ad esempio, il miagolio di un gatto, lo squillo di un telefono, lo schioccare di una frusta) è normale limita il difetto all'informazione uditiva verbale.

Questi pazienti hanno una memoria a lungo termine normale: la loro memoria per gli eventi della vita quotidiana, anche del passato remoto, è perfettamente conservata. Inoltre, riescono molto bene in tutti i compiti di laboratorio che verificano l'efficienza di tale forma di memoria, ad esempio nell'apprendimento di liste di parole o di un breve racconto. Mancano, infine, di difetti intellettuali generalizzati. Questi pazienti riescono spesso a condurre una vita pressoché normale. Il loro difetto può esser attribuito a una ridotta capacità del magazzino uditivo verbale a breve termine, che è in grado di contenere solo una quantità di informazione limitata in modo abnorme. L'ipotesi spiegherebbe il fatto che questi pazienti non sono in grado di ripetere correttamente più di due o tre stimoli verbali.

MAGAZZINO A BREVE TERMINE
MEMORIA OPERATIVA TEMPORANEA
STIMOLO AMBIENTALE
REGISTRI SENSORIALI
VISIVO
UDITIVO
TATTILE
MAGAZZINO A LUNGO TERMINE
MEMORIA PERMANENTE
PROCESSI DI CONTROLLO
RIPASSO DECISIONI
CODIFICAZIONE
STRATEGIE DI RICHIAMO
RISPOSTA

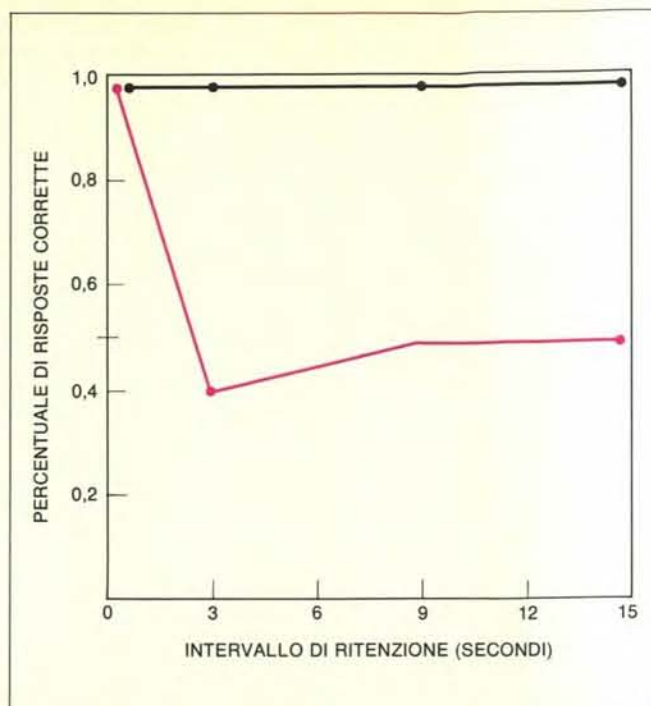
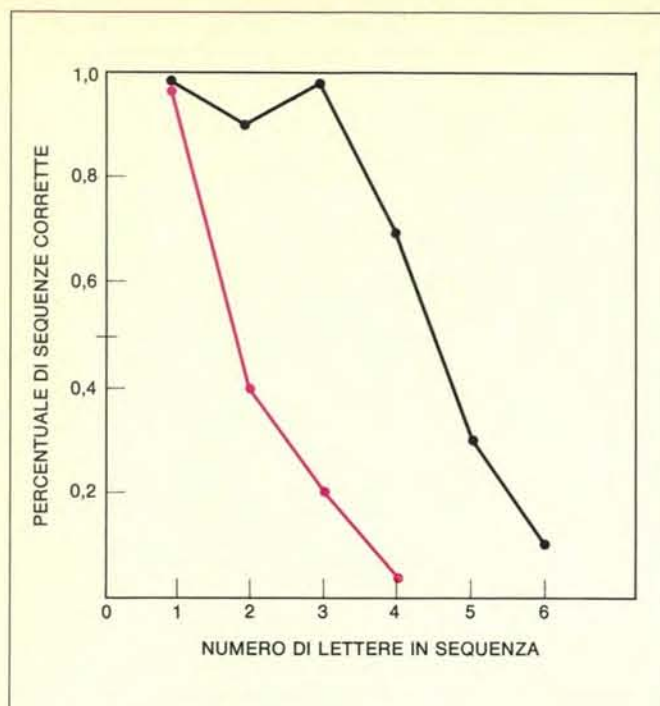
Secondo il modello degli psicologi americani Richard Atkinson e Richard Shiffrin della Stanford University in California, il flusso dell'informazione nei sistemi di memoria inizia con l'analisi dello stimolo ambientale. Poi l'informazione entra nel «magazzino a breve termine», che corrisponde alla memoria primaria di N. Waugh e D. Norman. Di qui può essere trasferita al «magazzino a lungo termine», che corrisponde alla memoria secondaria di Waugh e Norman, da dove può essere richiamata nel magazzino a breve termine. Ad esempio, quando viene visto un triangolo, il suo riconoscimento richiede il richiamo di tutte le informazioni relative a esso e contenute nel magazzino a lungo termine. I processi decisionali e di controllo della memoria a breve termine regolano il flusso dell'informazione al, e dal, magazzino a lungo termine, governando così l'apprendimento e l'oblio. Secondo questo modello, la memoria a breve termine ha una funzione cruciale nelle attività cognitive umane: permette di riconoscere gli stimoli ambientali, di apprendere nuove informazioni, di generare risposte. La memoria a breve termine è, nelle parole di Atkinson e Shiffrin, una «memoria operativa», «un sistema in cui vengono prese decisioni, vengono risolti problemi e viene diretto il flusso dell'informazione». Sia il modello relativamente semplice di Waugh e Norman, sia quello molto più complesso di Atkinson e Shiffrin, sono seriali: l'informazione, per essere stabilmente archiviata nella memoria a lungo termine, deve passare per quella a breve termine.

Il difetto selettivo della memoria a breve termine uditiva verbale è prodotto da lesioni, ad esempio infarti cerebrali, tumori o traumi, che interessano l'emisfero sinistro, in cui, nei soggetti destrimani, sono rappresentate le funzioni linguistiche. I casi finora descritti, tuttavia, sono ancora troppo pochi per poter localizzare con precisione all'interno dell'emisfero sinistro le regioni importanti per la memoria a breve termine. Secondo la neuropsicologa britannica Elizabeth Warrington del National Hospital for Nervous Diseases di Londra, un danno della porzione inferiore del lobo parietale sinistro è cruciale per produrre il difetto di memoria a breve termine.

La descrizione di questo tipo di disturbi cognitivi ha avuto un impatto profondo sulle concezioni dei processi di memoria normali. Certamente la distinzione tra memoria a breve termine e memoria a lungo termine ha trovato un'ulteriore conferma; tuttavia, i modelli a due fasi prima illustrati debbono essere ripensati in modo sostanziale, perché possano spiegare in modo adeguato il comportamento dei malati con difetto della memoria a breve termine uditiva verbale.

Si consideri il modello di Atkinson e Shiffrin, forse il più importante degli inizi degli anni settanta. Esso postula un solo magazzino a breve termine, a cui affluiscono informazioni dalle diverse modalità sensoriali. Una compromissione selettiva della memoria a breve termine uditiva verbale, associata a una migliore prestazione se lo stimolo è visivo, suggerisce che esistano magazzini a breve termine uditivi e visivi distinti. Sempre secondo lo stesso modello, la memoria a breve termine è qualcosa di più di un semplice magazzino in cui viene depositata l'informazione. È una «memoria operativa», in cui si possono distinguere: un magazzino a breve termine per contenere l'informazione e una serie di «processi di controllo», che dirigono l'analisi degli stimoli provenienti dal mondo esterno, il mantenimento dell'informazione nella memoria a breve termine e il suo eventuale trasferimento alla memoria a lungo termine, infine il richiamo dell'informazione archiviata nella memoria a lungo termine. Secondo il modello di Atkinson e Shiffrin, la memoria a breve termine ha funzioni molto importanti nell'ambito dei processi cognitivi e viene identificata con la cosciènza. L'esistenza di pazienti con un difetto selettivo della memoria a breve termine uditiva verbale non associato ad alcun altro importante disordine cognitivo induce a ritenere che la memoria a breve termine, intesa come «memoria operativa», non sia una struttura unitaria, ma possa a sua volta essere frazionata in diverse subcomponenti. Il tipo di difetto di questi pazienti può esser spiegato ammettendo un deficit del solo magazzino uditivo verbale, mentre i processi centrali di controllo sarebbero indenni. Appare chiaro, a questo punto, che la memoria a breve termine è molto più complessa di quanto il pur complicato modello di Atkinson e Shiffrin lasci supporre.

L'esistenza di pazienti con una ridotta capacità della memoria a breve termine uditiva verbale, non associata a una difettosa analisi fonologica dell'informazione, pone problemi anche a modelli come quello di Craik e Lockhart, che interpretano i processi di memoria in termini di livelli di analisi, ma non comprendono magazzini per l'informazione distinti dai processi di analisi. Secondo i due autori citati, si può prevedere che una lesione cerebrale produca un difetto a un livello di analisi (ad esempio fonologico o semantico), ma non una memoria a breve termine compromessa associata a una capacità di analisi indenne.

Il fatto che possa verificarsi una compromissione della memoria a breve termine senza danno della memoria a lungo termine impone, infine, una ridefinizione dei rapporti tra queste componenti della memoria. I modelli prima illustrati, con l'eccezione dell'impostazione per «livelli di analisi» di Craik e Lockhart, ammettono che l'informazione, per poter esser trasferita alla memoria a lungo termine, debba obbligatoriamente passare per la memoria a breve termine. L'esistenza di pazienti con un difetto della memoria a lungo termine può esser spiegata, ad esempio, ipotizzando un mancato trasferimento dell'informazione dal compartimento a breve termine a quello a lungo termine, oppure un'impossibilità a richiamare tale informazione dalla memoria a lungo termine. L'interpretazione del difetto di memoria a lungo termine è molto complessa. In questi ultimi anni tale forma di memoria è stata a sua volta frazionata in più componenti che, in analogia con quanto accade per la memoria a breve termine, possono esser selettivamente compromesse in pazienti con lesioni cerebrali. La trattazione di questi problemi esula dallo scopo di questo articolo. Ciò che importa osservare in questa sede è che i modelli a due componenti sopra discussi possono, in generale, spiegare un difetto della memoria a lungo termine associato a una memoria a breve termine indenne, ma non la dissociazione opposta. Se le due componenti della memoria sono in serie, non è spiegabile un difetto della memoria a breve ter-

Il difetto più evidente dei pazienti con danno della memoria a breve termine verbale è l'incapacità di ripetere, subito dopo la presentazione, sequenze di stimoli uditivi verbali. Viene mostrato (*figura a sinistra*) il comportamento di una paziente con difetto della memoria a breve termine uditiva verbale, non associata ad altri disordini afasici. La lesione è di natura ischemica e interessa le regioni peri-silviane dell'emisfero sinistro. Quando lo stimolo è uditivo (*curva in colore*), la paziente è in grado di ripetere solo sequenze costituite da una sola lettera (ad esempio, *C*); già quando le lettere sono due (ad esempio, *FZ*) vi è un marcato deterioramento della prestazione e la difficoltà è ancora maggiore per sequenze di tre o quattro lettere (ad esempio, *GRS* o *QZRD*). Quando, invece, l'informazione è presentata nella modalità visiva (*curva in nero*), la prestazione è nettamente migliore. Questa dissociazione suggerisce l'esistenza di un magazzino a breve termine visivo, separato da quello uditivo. Viene anche illustrato (*figura a destra*) il comportamento della stessa paziente in un compito in cui è richiesto il ricordo di stimoli verbali, presentati acusticamente o visivamente. Dopo la presentazione, la paziente viene distratta e impegnata in un compito aritmetico per un periodo di tempo variabile tra tre e quindici secondi. Al termine di questo periodo le viene chiesto di ripetere le lettere che le erano state presentate. Nei soggetti normali si osserva un oblio significativo nel giro di alcuni secondi se l'informazione da ricordare consiste in digrammi o in trigrammi (ad esempio, *TS* o *PST*), senza differenze rilevanti in relazione alla modalità acustica o visiva di presentazione dello stimolo. I pazienti in cui la memoria a breve termine uditiva verbale è difettosa hanno, invece, un oblio marcato, anche se viene loro chiesto di ricordare una singola lettera, qualora la presentazione sia uditiva. La figura mostra che, nella paziente sopra citata, vi è un oblio importante già dopo tre secondi di distrazione, anche se lo stimolo è costituito da una sola lettera (*curva in colore*). La specificità del difetto è dimostrata dal fatto che quando lo stimolo è presentato visivamente (*curva in nero*) non vi è oblio, come accade in un soggetto normale. Inoltre, l'osservazione che la ripetizione immediata - a zero secondi di intervallo dalla presentazione - è perfetta esclude la possibilità che il difetto possa essere interpretato in termini di difettosa percezione dello stimolo.

mine associato a una memoria a lungo termine normale. Si rende quindi necessario postulare che siano possibili accessi diretti e indipendenti ai due sistemi di memoria.

Appare chiaro, dunque, come gli studi di pazienti portatori di lesioni cerebrali abbiano avuto un forte impatto sulle concezioni dei normali processi di memoria, rendendo necessarie sostanziali modifiche dei modelli psicologici che cercano di rappresentare la memoria umana.

È invece meno evidente a questo punto a che cosa serva la memoria a breve termine. Quella a lungo termine ha un'utilità ovvia: da un lato permette di ricordare ciò che abbiamo acquisito nel passato, dall'altro ci permette di apprendere e ritenere nuove informazioni in maniera stabile. Per quanto concerne, invece, la memoria a breve termine, le tecniche con le quali essa viene comunemente studiata - ad esempio la ripetizione di serie di stimoli verbali, per lo più senza senso - non danno alcun suggerimento relativo alla sua utilità. Inoltre, i pazienti con danno della memoria a breve termine uditiva verbale, eccezion fatta per la difficoltà a ripetere immediatamente sequenze di stimoli verbali, non sembrano avere altri importanti deficit cognitivi e spesso conducono una vita normale, svolgendo anche attività professionali impegnative. Parrebbe così che la memoria a breve termine verbale sia utile solo in situazioni alquanto improbabili: ad esempio, per tenere a mente un numero telefonico che ci è stato appena detto, se il telefono ove dobbiamo comporre il numero si trova in un'altra stanza e non abbiamo carta e penna a disposizione per scriverlo. È certamente poco, solo se si pensa che lo studio di questa memoria ha tenuto impegnati per molti anni psicologi sperimentali di prim'ordine.

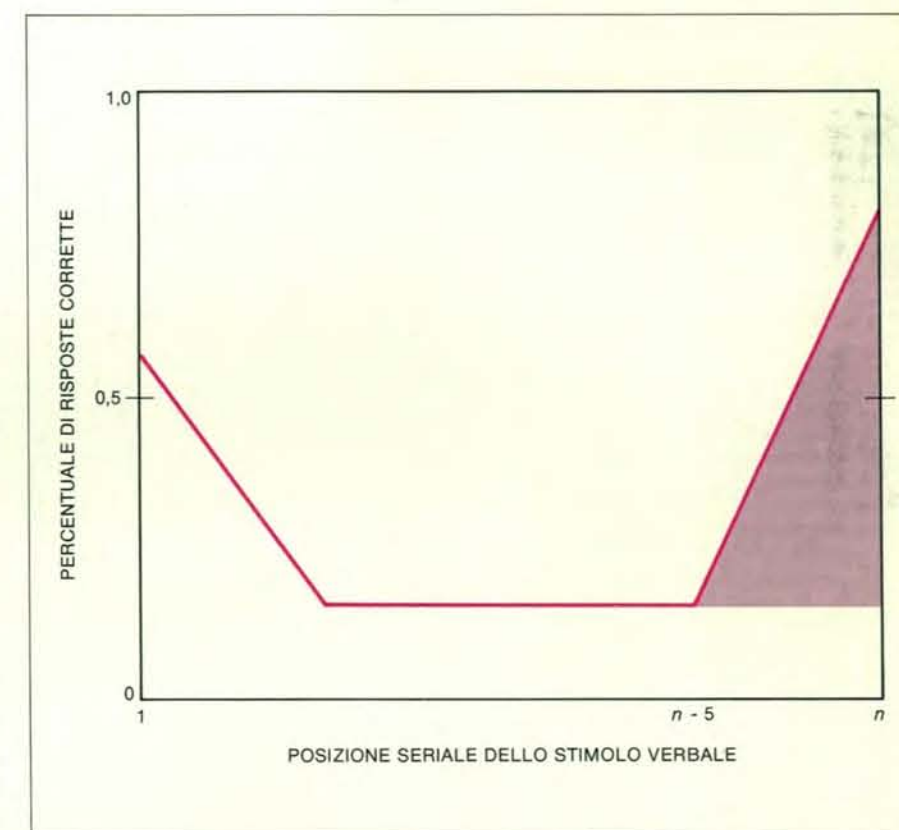
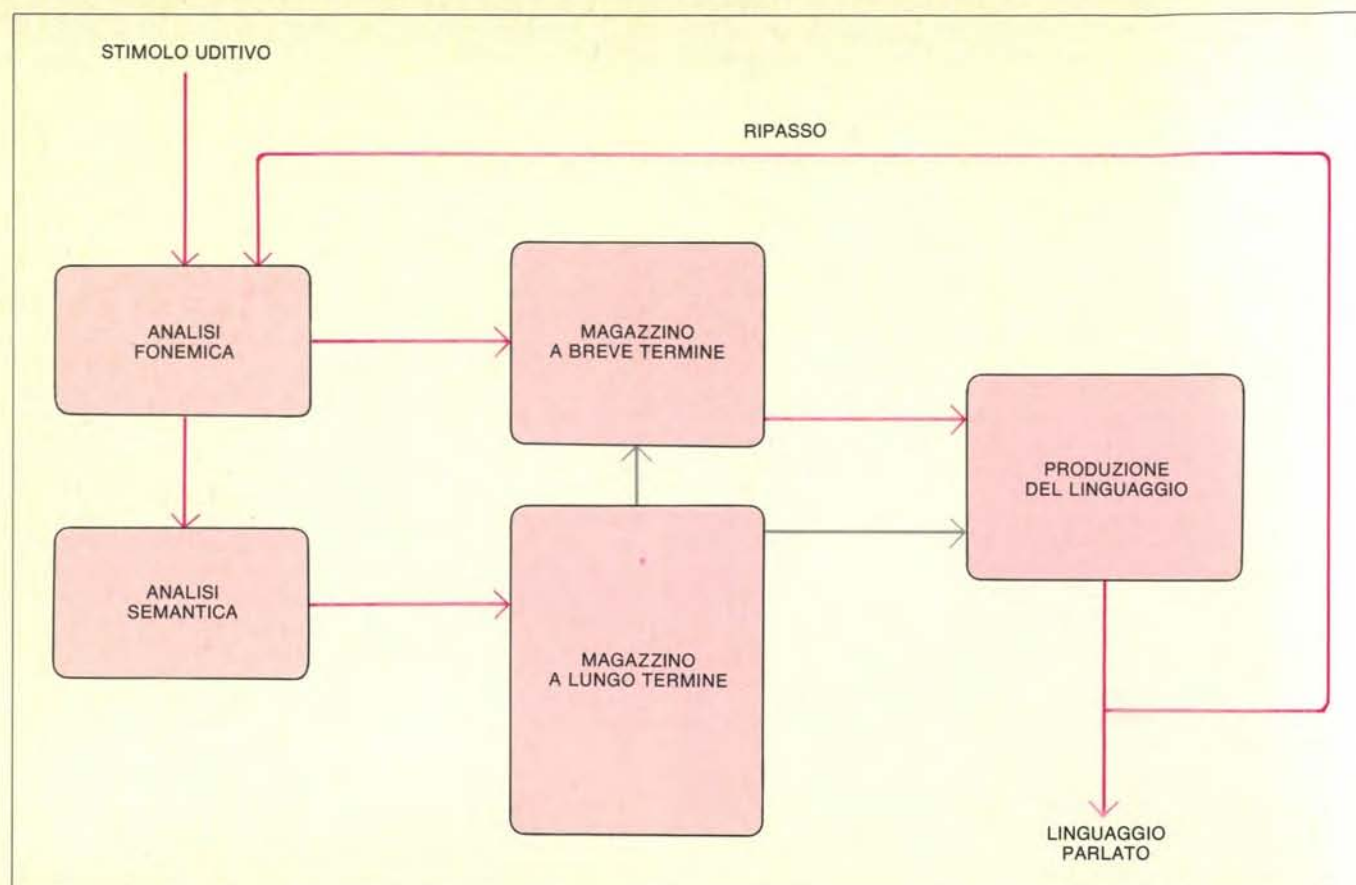
Lo studio del comportamento di malati con difetto di memoria a breve termine verbale ha consentito di dimostrare in modo convincente che tale memoria è qualcosa di più del prodotto di esperimenti di laboratorio. Questi pazienti, infatti, presentano difetti di comprensione molto sottili e molto particolari.

In una normale conversazione, la comprensione del discorso di un paziente con una memoria a breve termine uditiva verbale difettosa appare più o meno normale. Un'analisi più dettagliata della comprensione del linguaggio mostra, invece, un quadro che può esser così sintetizzato. La comprensione di parole isolate è normale: ad esempio, se al paziente viene presentata la parola «cane» e gli viene chiesto di indicare la figura corrispondente fra alcune alternative, egli è in grado di eseguire bene tale compito. Anche la comprensione di frasi semplici (ad esempio, «Gli architetti progettano edifici»), ove la comprensione dei principali elementi lessicali è sufficiente ad assicurare una buona prestazione, è conservata. Il rendimento di questi pazienti si deteriora, invece, per frasi di maggiore complessità. Ad esempio, un test molto usato per diagnosticare disturbi del linguaggio di tipo afasico, il «test dei gettoni» proposto nel 1962 dai neuropsicologi Ennio De Renzi e Luigi Amedeo Vignolo della Clinica di malattie nervose e mentali dell'Università di Milano (si veda l'articolo *Intelligenza e afasia* di Anna Basso, in «Le Scienze», n. 145, settembre 1980), richiede al soggetto di manipolare gettoni di varia forma e colore in risposta a ordini non ridondanti di variabile complessità. I pazienti con difetto di memoria a breve termine uditiva verbale eseguono bene ordini relativamente semplici (ad esempio, «Tocchi il gettone rosso», oppure «Tocchi il quadrato giallo piccolo»). Commettono, invece, errori quando gli ordini sono più lunghi e complessi (ad esempio, «Tocchi il quadrato verde piccolo e il cerchio rosso grande»). In un caso di difetto selettivo della memoria a breve termine uditiva verbale, studiato da Anna Basso, Hans Spinnler, Maria Ester Zanobio e dall'autore, tutti dell'Università di Milano, si è osservato che la comprensione degli ordini impartiti nel test sopra menzionato è migliore se essi vengono presentati nella modalità visiva, come ci si potrebbe del resto attendere in questo tipo di pazienti. Anche la comprensione di frasi complesse, il cui significato è legato all'ordine degli elementi verbali, è compromessa (ad esempio, «L'uomo che il bambino colpì portava una scatola» può divenire «L'uomo colpì il bambino che portava una scatola»).

Come è possibile interpretare questi difetti di comprensione? Si può pensare che una frase, quando è lunga e complessa (e l'ordine degli elementi verbali è cruciale per intender correttamente il suo significato) non può essere compresa dall'ascoltatore a mano a mano che la percepisce. Deve, invece, esser mantenuta in una memoria operativa il tempo necessario perché le procedure necessarie alla sua comprensione vengano espletate. Supponiamo che, a seguito di una lesione cerebrale, la memoria a breve termine uditiva verbale sia lesa. La comprensione uditiva di singoli elementi lessicali e di frasi semplici e corte sarà relativamente conservata, in quanto il carico imposto alla memoria a breve termine sarà minimo o nullo; al contrario, gli ordini più lunghi e complessi che fanno parte del test dei gettoni, e le frasi complesse semanticamente reversibili di cui si è dato esempio, non possono esser probabilmente compresi durante la presentazione: richiedono, invece, di esser temporaneamente mantenuti in una memoria operativa, che consenta la conservazione della struttura superficiale della frase (gli elementi costitutivi nella loro disposizione lineare). In una situazione di questo tipo la comprensione dei pazienti con difetto della memoria a breve termine verbale si deteriora.

Le ricerche fin qui discusse suggeriscono che la memoria a breve termine uditiva verbale sia importante per la comprensione del linguaggio, fungendo da magazzino temporaneo per frasi lunghe e complesse, il cui significato non può esser compreso dal soggetto a mano a mano che le ascolta. Secondo questa interpretazione, e secondo i modelli che ho fin qui preso in considerazione, la memoria a breve termine uditiva verbale ha una collocazione «in input» piuttosto che «in output»: cioè è interessata nella ritenzione dell'informazione verbale che giunge al soggetto, ma non nella produzione del linguaggio parlato.

La produzione del linguaggio è un processo molto complesso, che va dalla pianificazione dell'intero discorso e delle frasi che lo compongono alla messa in opera di un programma articolatorio che specifichi quali muscoli debbano entrare in funzione e in quale sequenza temporale. Secondo alcuni ricercatori, tra i quali lo psicologo britannico Andrew Ellis dell'Università di Lancaster, tra le diverse componenti funzionali che concorrono alla produzione di un eloquio fluente, è ipotizzabile l'esistenza di un magazzino a breve termine a capacità limitata, con locus «in output», che conterrebbe sequenze preprogrammate di parole che stanno per esser prodotte. Tale componente funzionale, nella quale sequenze di alcune parole sarebbero codificate in una forma articolatoria, può spiegare, ad esempio, il fenomeno della «coarticolazione», cioè il fatto che, durante l'eloquio, le strutture anatomiche deputate all'articolazione (la bocca, la lingua, i denti) assumono posture determinate anche dalle consonanti e dalle vocali che non sono state ancora prodotte e non solo da quelle articolate in quel momento.

Anche alcuni tipi di *lapsus linguae* possono essere interpretati come un argomento a favore dell'esistenza di un magazzino a breve termine «in output». È il caso del fenomeno dello «spoonerismo». Si tramanda che l'ecclesiastico inglese William Spooner, vissuto tra la fine dell'Ottocento e l'inizio del Novecento, commettesse, nel parlare, molti lapsus in cui l'ordine seriale delle consonanti o delle vocali veniva alterato: ad esempio, «you have *m*issed all my *h*istory lectures» («avete mancato a tutte le mie conferenze di storia) diveniva «you have *h*issed all my

La tecnica della rievocazione libera di liste di parole è stata molto utilizzata in psicologia normale per riuscire a dissociare le due componenti della memoria. In base a essa, vengono presentate al soggetto liste costituite da dieci o più parole, ad esempio: cane, albero... rosa, muro. Subito dopo la presentazione, il soggetto ripete tutte le parole che riesce a ricordare, nell'ordine che preferisce. Esperimenti di questo tipo mostrano invariabilmente che le ultime parole e le prime parole della lista sono meglio ricordate delle parole che si trovano al centro della lista stessa. Questi fenomeni, chiamati rispettivamente effetti «di ultima posizione» e «di prima posizione», danno alla curva di rievocazione un tipico aspetto a *U*. L'effetto di ultima posizione è transitorio: qualora la rievocazione sia differita di trenta secondi, esso infatti scompare. L'effetto di prima posizione è comunemente attribuito alla memoria a lungo termine, mentre la memoria a breve termine sarebbe responsabile dell'effetto di ultima posizione. In altri termini, le ultime parole ricordate della lista proverrebbero dalla memoria a breve termine, le prime dalla memoria a lungo termine. In accordo con questa osservazione è stato rilevato che, nei pazienti che presentano un difetto della memoria a breve termine uditiva verbale, l'effetto di ultima posizione è molto ridotto o assente, mentre l'effetto di prima posizione si conserva. Nei pazienti amnesici con difetto della memoria a lungo termine è stata osservata, invece, la dissociazione opposta: mentre l'effetto di prima posizione scompare, quello di ultima posizione è presente.

Secondo i neuropsicologi britannici Tim Shallice ed Elizabeth Warrington, i risultati degli esperimenti che vengono condotti sui pazienti che presentano un difetto selettivo della memoria a breve termine uditiva verbale suggeriscono che l'informazione possa accedere direttamente al magazzino a lungo termine, senza dover passare in precedenza per il magazzino a breve termine. I due magazzini di memoria in questo modello sono posti «in parallelo» e non «in serie». Le connessioni della memoria a lungo termine sono lasciate volutamente imprecisate (*frecce in grigio*) in quanto non sono state esaminate direttamente negli esperimenti condotti dai due autori del modello.

*my*stery lectures» (avete fischiato tutte le mie conferenze sul mistero). Non è affatto certo che il reverendo Spooner abbia commesso tutti i lapsus, spesso spiritosi come quello citato, a lui attribuiti. Si dice, inoltre, che spesso li commettesse deliberatamente. A ogni modo il termine «spoonerismo fonemico» si riferisce a lapsus in cui vi è un'alterazione dell'ordine di vocali o consonanti. Anche Sigmund Freud, nel suo libro *Psicopatologia della vita quotidiana,* cita lapsus di questo tipo («mi sentivo il pe*ss*o... pe*tt*o oppre*ss*o»). Sia il fenomeno della coarticolazione sia lo spoonerismo fonemico possono esser spiegati ammettendo l'esistenza di una programmazione fonemica e di un magazzino che contenga sequenze preprogrammate di linguaggio, che stanno per esser articolate. Secondo Ellis, tale componente coincide con la memoria a breve termine che - come abbiamo visto - è cruciale per il ricordo immediato di quantità limitate di materiale verbale e per certi aspetti della comprensione del linguaggio. Un argomento a favore di questa tesi viene dall'osservazione che i soggetti normali compiono, nei compiti di memoria a breve termine, errori molto simili a certi errori del linguaggio spontaneo. Quest'affermazione risulta più comprensibile se si considera che lettere che hanno un suono simile (ad esempio *B, P, T, V, D*) si assomigliano anche per come vengono articolate. Quindi, gli errori di confusione «acustica» caratteristici della memoria a breve termine potrebbero essere descritti come errori di confusione «articolatoria». E, in effetti, alcuni studi condotti su soggetti sani sono a favore di una rappresentazione «articolatoria» dell'informazione verbale nella memoria a breve termine.

Il problema del locus «in input» o «in output» della memoria a breve termine verbale è - come si vede - piuttosto complesso. Lo studio dei pazienti con deficit della memoria a breve termine ha dato in proposito qualche indicazione. In alcuni casi, ad esempio nel caso del paziente K. F. di Warrington e Shallice, il difetto della memoria a breve termine era associato ad alterazioni dell'espressione: un eloquio lento ed esitante, difficoltà a trovar la parola appropriata e parafasie soprattutto fonemiche (errori di sostituzione di fonemi all'interno delle parole: ad esempio, *t*e*v*olo al posto di *t*a*v*olo). Un caso di questo genere corrobora l'ipotesi che un'unica componente funzionale a breve termine sia impegnata e nella memoria verbale e nella produzione di un linguaggio fluente.

In altri pazienti, invece, ad esempio nel caso J. B. studiato da Shallice e dallo psicologo Brian Butterworth dell'Università di Cambridge e nel caso P. V. da noi osservato, il difetto della memoria a breve termine uditiva verbale non era associato ad anomalie espressive di rilievo.

Che cosa suggeriscono questi dati neuropsicologici? L'osservazione che, almeno in taluni casi, il deficit della memoria a breve termine verbale non si associa a un disordine della produzione del linguaggio farebbe pensare che la componente funzionale impegnata in compiti quali la ripetizione immediata di sequenze di cifre non sia la stessa interessata nella produzione del linguaggio. Il fatto che taluni pazienti presentino anche disturbi della produzione del linguaggio può esser addebitato al fatto che le lesioni cerebrali (infarti cerebrali, emorragie, tumori o traumi) spesso compromettono più di un sistema funzionale. Di qui discende l'importanza dei casi cosiddetti «puri», con difetti funzionali molto selettivi. Lo studio di questi pazienti, purtroppo assai rari, costituisce un utile metodo per validare, o falsificare, modelli delle attività cognitive normali.

Reattori a fusione a confinamento magnetico

I progetti attualmente in corso e i reattori sperimentali in costruzione consentiranno di verificare la fattibilità della produzione di energia mediante la fusione termonucleare di ioni confinati da campi magnetici

di Robert W. Conn

La ricerca relativa allo sfruttamento dell'energia da fusione nucleare con l'impiego di campi magnetici per il confinamento di un plasma ad altissima temperatura (un plasma è un gas elettricamente neutro, costituito da particelle cariche) e la connessa produzione di energia elettrica è in continuo sviluppo dall'inizio degli anni cinquanta e si sta ora apprestando a fornire risultati di interesse pratico. Con un costo di costruzione di 314 milioni di dollari, gli Stati Uniti hanno realizzato l'apparecchiatura toroidale Tokamak Fusion Test Reactor (TFTR), entrata in funzione nel dicembre 1982 al Plasma Physics Laboratory di Princeton, nel New Jersey. La Comunità economica europea ha realizzato e recentemente avviato il Joint European Torus, noto anche come tokamak JET, al Culham Laboratory, presso Oxford. Un'apparecchiatura analoga, che dovrebbe funzionare nel 1985, il tokamak JT-60, è in costruzione in Giappone. L'insieme di questi tre reattori rappresenta un investimento di oltre due miliardi di dollari. I programmi sperimentali del TFTR e del JET culmineranno nel 1986, con la «combustione» di una miscela di deuterio e trizio. (Deuterio e trizio sono gli isotopi pesanti dell'idrogeno: H-2 e H-3; costituiranno inoltre il combustibile dei reattori a fusione di potenza, ossia dei reattori a fusione di tipo commerciale.) Il JT-60 consumerà solo deuterio e l'abbondante isotopo leggero dell'idrogeno: H-1.

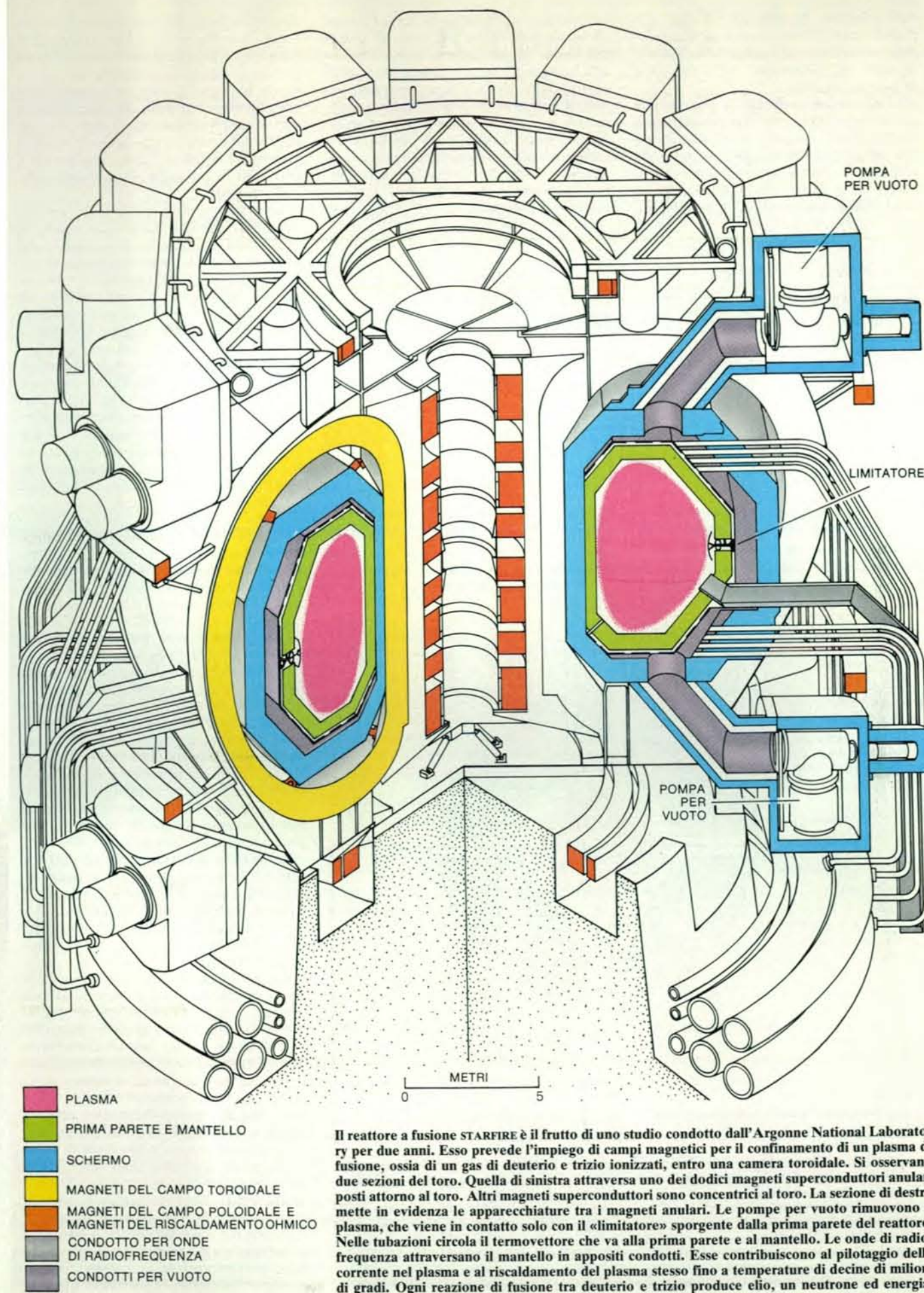
È inoltre allo studio una soluzione alternativa a quella del tokamak. Negli Stati Uniti, al Lawrence Livermore National Laboratory, in California, è in costruzione la Mirror Fusion Test Facility B (MFTF-B). Si tratta di uno specchio tandem, un dispositivo in cui un plasma è confinato da barriere magnetiche ed elettrostatiche a ciascuna estremità di una sequenza lineare di magneti. Il suo avvio è previsto per il 1986, con un costo di costruzione previsto di 209 milioni di dollari. Obiettivo dell'MFTF-B è il confinamento di un plasma di deuterio in condizioni prossime a quelle corrispondenti a un reattore di potenza a specchio tandem. L'MFTF-B consentirà così di verificare la fattibilità scientifica della soluzione a specchio tandem.

L'MFTF-B viene costruito con magneti superconduttori al fine di minimizzare il suo consumo di energia elettrica. È già stata realizzata e provata con esito positivo una importante serie di bobine costituenti una coppia di magneti a forma di lettera C, detta configurazione «yin-yang». Altri due tokamak di grandi dimensioni, il *T*-15 in Unione Sovietica, la cui costruzione sarà prevedibilmente ultimata nel 1986, e il TORE-SUPRA in Francia, che sarà completato entro il 1988, saranno pure equipaggiati con magneti superconduttori.

La costruzione di queste grosse macchine sperimentali venne decisa negli anni settanta sulla base di tre considerazioni fondamentali. Innanzitutto, le ricerche sul confinamento magnetico avevano prodotto una serie di risultati di rilievo (e continuano a farlo). In secondo luogo, eventi quali l'embargo del 1973 imposto dall'Organizzazione dei paesi esportatori di petrolio (OPEC) avevano posto in evidenza la vulnerabilità del mondo industrializzato alla interruzione delle forniture energetiche. In terzo luogo, aveva preso corpo la consapevolezza delle possibili limitazioni nelle quantità e negli impieghi dei combustibili fossili. Gli esperimenti di fusione si svolgeranno dunque negli anni ottanta e, se tutto andrà come previsto, sarà dimostrata la fattibilità scientifica di un reattore a fusione a confinamento magnetico. Rimarrà comunque l'interrogativo se la fusione magnetica sia fattibile anche in termini di ingegneria, economia e sicurezza.

La reazione di fusione tra deuterio e trizio comporta il rispetto di una condizione rigorosa: il superamento della repulsione elettrostatica tra i nuclei di deuterio e trizio (deutoni e tritoni), entrambi dotati di una carica unitaria positiva, mediante l'innalzamento della loro velocità relativa fino a raggiungere un valore corrispondente a un'energia di 100 000 elettronvolt (100 keV). Un gas di deuterio e trizio presenta un'ampia distribuzione di energie e quindi al mantenimento della combustione termonucleare da parte del gas bastano, in media, 5 keV, che corrispondono, però, a una temperatura di 58 milioni di gradi centigradi. A tale temperatura gli elettroni non sono legati ai nuclei e danno luogo al plasma di fusione.

Ogni reazione di fusione produce un nucleo di elio 4 (una particella alfa) di alta energia e un neutrone molto veloce. Essendo privo di carica elettrica, il neutrone sfugge dal plasma. La particella alfa non sfugge, in quanto la sua doppia carica positiva unitaria la tiene vincolata allo stesso campo magnetico impiegato per il confinamento del plasma. La particella alfa ha un'energia di 3520 keV, molto superiore a quella delle particelle del plasma, cosicché, rallentando in seguito agli urti, cede la propria energia in eccesso ai deutoni e ai tritoni, contribuendo a mantenerne elevata la temperatura.

La condizione necessaria perché il plasma sia in grado di mantenere la propria combustione termonucleare - ossia la condizione di ignizione - è che il «parametro di confinamento», dato dal prodotto della densità del plasma (in particelle per centimetro cubo) per il tempo medio di perdita di energia (espresso in secondi), superi il valore di 3×10^{14}. Se il parametro di confinamento è inferiore a quel valore, non si può mantenere la temperatura del plasma senza ricorrere al riscaldamento ausiliario. Il plasma può però produrre energia, comportandosi in pratica come un amplificatore ad alto guadagno della potenza ausiliaria. La misura del guadagno è data dal valore Q, che rappresenta il rapporto tra la potenza di fusione generata e la potenza immessa per sostenere la temperatura del plasma. All'atto

Il reattore a fusione STARFIRE è il frutto di uno studio condotto dall'Argonne National Laboratory per due anni. Esso prevede l'impiego di campi magnetici per il confinamento di un plasma di fusione, ossia di un gas di deuterio e trizio ionizzati, entro una camera toroidale. Si osservano due sezioni del toro. Quella di sinistra attraversa uno dei dodici magneti superconduttori anulari posti attorno al toro. Altri magneti superconduttori sono concentrici al toro. La sezione di destra mette in evidenza le apparecchiature tra i magneti anulari. Le pompe per vuoto rimuovono il plasma, che viene in contatto solo con il «limitatore» sporgente dalla prima parete del reattore. Nelle tubazioni circola il termovettore che va alla prima parete e al mantello. Le onde di radiofrequenza attraversano il mantello in appositi condotti. Esse contribuiscono al pilotaggio della corrente nel plasma e al riscaldamento del plasma stesso fino a temperature di decine di milioni di gradi. Ogni reazione di fusione tra deuterio e trizio produce elio, un neutrone ed energia.

dell'ignizione, Q diventa infinito. Un reattore con Q finito è una macchina pilotata, perché in esso l'entità della fusione dipende dalle variazioni nell'immissione di potenza ausiliaria.

Come viene confinato il plasma? La classificazione più generale dei sistemi di confinamento magnetico distingue tra due ampie classi di sistemi a seconda che le linee di forza del campo magnetico siano chiuse o aperte. Esempio classico di linee chiuse sono quelle che definiscono un volume toroidale: in questo caso ogni particella del plasma è vincolata a una di tali linee. In particolare, ruota attorno alla linea pur potendo liberamente percorrerla, descrivendo così una traiettoria a elica. La particella è vincolata in quanto la linea si chiude su se stessa. Nel progetto iniziale detto «stellarator», le linee del campo sono create unicamente da avvolgimenti esterni al plasma. Una volta ignito, il plasma resta tale senza apporto esterno di energia cosicché il guadagno Q diventa infinito. In un progetto successivo, il tokamak, le linee sono prodotte in parte da avvolgimenti esterni e in parte da una corrente elettrica indotta nel plasma stesso. In pratica, alcuni avvolgimenti dello stellarator sono sostituiti da una nuova serie di avvolgimenti che fungono da avvolgimento primario di un trasformatore a corrente continua, il cui secondario è costituito da un unico avvolgimento: l'anello di plasma.

TRIZIO
DEUTERIO
GENERATORE DI VAPORE E TURBINA
SCAMBIATORI DI CALORE SECONDARI
SCAMBIATORI DI CALORE PRIMARI
ESTRAZIONE DEL TRIZIO
TERMOVETTORE
RECUPERO DEL COMBUSTIBILE, RITRATTAMENTO, DEPOSITO E RICICLO
RIMOZIONE DELLE «CENERI» E POMPAGGIO PER VUOTO
PRIMA PARETE
RISCALDAMENTO AUSILIARIO
PLASMA
MANTELLO
SCHERMO
MAGNETI
INIEZIONE DEL COMBUSTIBILE

Questo disegno schematico presenta gli aspetti ingegneristici comuni a tutti i reattori a fusione a confinamento magnetico. Un campo magnetico confina il plasma di fusione; un sistema di riscaldamento ausiliario contribuisce a elevarne la temperatura e un sistema di riciclo del combustibile provvede al mantenimento della purezza del plasma e alla sua alimentazione con combustibile termonucleare. La prima parete del reattore deve resistere al calore emesso dal plasma. I neutroni emessi dal plasma devono penetrare nel mantello e l'energia che vi depositano viene trasmessa (sotto forma di calore) a un termovettore. A sua volta il termovettore genera vapore, che aziona le turbine per la generazione di energia elettrica. Attraverso reazioni nucleari, nel mantello avviene la fertilizzazione del trizio, un isotopo radioattivo che è estremamente raro in natura.

Un trasformatore non è però in grado di pilotare con continuità la corrente nel secondario in un'unica direzione perché per fare ciò la corrente nel primario dovrebbe crescere indefinitamente. Occorre quindi ripristinare il trasformatore in modo che la combustione termonucleare possa cessare periodicamente. Secondo le stime attuali, i reattori tokamak produrranno energia in impulsi della durata di circa 1000 secondi, intervallati da pause di 30-60 secondi. Nei reattori commerciali, un sistema di accumulo di energia consentirebbe l'erogazione continua di potenza alle turbine elettrogeneratrici, ma le pause imporrebbero limiti progettuali ai componenti del reattore per il connesso incremento della fatica ciclica.

Sono perciò allo studio altri metodi di pilotaggio della corrente del plasma. Se avranno successo, essi permetteranno il funzionamento stazionario dei tokamak, o per lo meno il funzionamento a impulsi di lunghissima durata, dell'ordine di un giorno. L'idea più promettente è quella di inviare onde di radiofrequenza (RF) monodirezionali nel plasma a una frequenza di circa tre miliardi di hertz. La quantità di moto delle onde può trasmettersi agli elettroni, sospingendoli lungo le linee del campo magnetico con conseguente generazione di corrente elettrica. La fattibilità del processo è stata dimostrata in esperimenti svoltisi al Massachusetts Institute of Technology e alla Princeton University. Resta da vedere se tale processo sarà efficace ai valori di densità di un plasma di fusione. Per un reattore con pilotaggio di corrente a radiofrequenza il valore di Q fornito dalle stime teoriche è compreso tra 10 e 20. Un valore di Q uguale a 20 è in pratica il minimo accettabile per un reattore tokamak commerciale.

La seconda classe generale di sistemi di confinamento magnetico prevede linee di campo aperte. La configurazione fondamentale è data da uno specchio magnetico a cella singola: in questo caso una serie di magneti produce un campo la cui intensità presenta due picchi. Ciascun picco rappresenta una zona a «collo di bottiglia», ove le linee di forza sono strettamente addensate. Si consideri una particella carica vincolata a una linea di forza in un punto qualsiasi fra i due picchi. Mentre la particella si avvicina all'uno o all'altro picco deve sempre conservare la propria energia e il proprio momento angolare; il moto della particella parallelo al campo viene quindi trasformato in moto perpendicolare allo stesso, con la conseguenza che la particella viene in pra-

Lo STARFIRE adotta il sistema di confinamento del plasma di tipo tokamak, che utilizza tre gruppi di magneti. I magneti del campo toroidale generano un campo le cui linee di forza sono concentriche al toro formato dal plasma. I magneti di riscaldamento ohmico hanno una funzione analoga a quella degli avvolgimenti del primario di un trasformatore, inducendo una corrente elettrica nell'avvolgimento secondario, che è costituito dal plasma stesso. La corrente provoca il riscaldamento del plasma; inoltre, genera un campo poloidale avente linee di forza anulari disposte perpendicolarmente al campo toroidale. I magneti del campo poloidale, oltre a contribuire all'efficacia del trasformatore, generano un campo con linee di forza verticali che impedisce la fuga del plasma dalla geometria toroidale. Le linee di forza del campo risultante sono elicoidali. Ogni particella carica nel plasma ruota attorno a una linea di forza e progredisce mentre la percorre; la particella risulta vincolata perché la linea è chiusa (cioè dopo alcuni giri attorno alla colonna toroidale la particella si ritrova al punto di partenza).

tica a trovarsi confinata tra i due picchi.

Uno specchio magnetico a cella singola non è un contenitore perfetto: infatti le particelle, urtandosi e diffondendo, possono sfuggire dalla strozzatura del collo di bottiglia. Gli elettroni diffondono molto più rapidamente degli ioni cosicché sfuggono per primi. Ne consegue che il plasma restante sviluppa un potenziale elettrostatico positivo, che riduce l'ulteriore flusso di elettroni della quantità sufficiente a rendere equivalenti le correnti di elettroni e di ioni. Per mantenere la densità degli ioni si possono iniettare nel plasma fasci di atomi neutri con un'energia intorno a 200 keV. Tali atomi attraversano inalterati il campo magnetico e vengono ionizzati e confinati quando urtano gli elettroni del plasma. Tuttavia, anche con l'adozione di questa tecnica, il valore di Q calcolato per uno specchio magnetico a cella singola non supera l'unità.

In alternativa, si possono collegare magneticamente due celle-specchio, formando così uno specchio tandem. In questo caso, il picco di potenziale elettrostatico in ciascuna cella agisce da trappola elettrostatica degli ioni, che vengono confinati in una valle centrale di potenziale tra le celle-specchio. Nello stesso tempo una valle più ridotta ma più profonda, all'interno di ciascuna cella-specchio, isola gli elettroni delle celle da quelli della valle centrale. Questa barriera consente di portare ad altissima energia gli elettroni delle celle-specchio e quindi di creare i potenziali di tali celle con grande efficienza. Gli elettroni vengono riscaldati mediante onde di radiofrequenza, la cui frequenza, tipicamente compresa tra 30 e 100 miliardi di hertz, corrisponde alla frequenza di ciclotrone dell'elettrone (la frequenza, cioè, del moto a spirale dell'elettrone attorno alle linee del campo magnetico del plasma). La fattibilità del riscaldamento degli elettroni delle celle-specchio terminali e del confinamento degli ioni della cella centrale è stata recentemente dimostrata al Lawrence Livermore Laboratory.

Per sostenere il plasma delle celle-specchio in uno specchio tandem occorre fornire continuamente energia; si valuta che Q possa variare tra 10 e 30. È però importante osservare che l'energia delle particelle cariche, perse dalle estremità del reattore, può essere recuperata indirizzando le particelle in apparecchiature che funzionano sostanzialmente al contrario degli acceleratori di particelle. La generazione di energia elettrica, supplementare a quella derivante dalle reazioni di fusione, mediante questi «convertitori diretti» può avere un rendimento superiore del 50 per cento rispetto alla generazione di energia elettrica mediante riscaldamento di acqua e azionamento di turbine con il vapore prodotto. Si valuta che, per un reattore di potenza a specchio tandem, il valore di Q debba essere intorno a 10, circa la metà quindi del valore minimo per un corrispondente reattore tokamak.

I primi consistenti studi sulla fattibilità dei reattori a fusione risalgono all'inizio degli anni settanta. Almeno due ne erano gli obiettivi di fondo. In primo luogo, determinare la portata dei problemi ingegneristici più complessi; sotto questo aspetto i risultati hanno spesso condotto a ulteriori ricerche specifiche. In secondo luogo, mettere in evidenza che taluni aspetti della fisica del reattore sono in grado di pregiudicare la fattibilità ingegneristica di particolari sistemi di confinamento; sotto quest'altro aspetto, i risultati ottenuti hanno spesso favorito svilup-

pi concettuali in fisica che hanno contribuito ad attenuare la complessità dei problemi di ingegneria.

Gli aspetti fondamentali dell'ingegneria dei reattori a fusione a confinamento magnetico sono comuni a tutti i sistemi di confinamento. In primo luogo, la maggior parte dei progetti di tali reattori prevede l'impiego di magneti superconduttori per ridurre al minimo la quantità di energia necessaria al mantenimento del campo magnetico. Per un volume compreso tra 3000 e 10 000 metri cubi il valore di riferimento per l'intensità del campo magnetico è di 50 000 gauss. (Per confronto, il valore medio dell'intensità del campo magnetico terrestre è di circa 0,5 gauss.) Se si adottassero magneti non superconduttori, l'energia assorbita dai magneti supererebbe quella prodotta dal reattore.

La superconduttività è una proprietà comune a molti metalli e leghe: in sostanza, se tali materiali vengono raffreddati al di sotto di una certa temperatura critica, danno luogo a una transizione di fase che ne azzera la resistività elettrica. Nelle leghe reali, quali niobio-titanio in rapporto uno a uno o niobio-stagno in rapporto tre a uno, la superconduttività si manifesta a temperature inferiori ai 20 kelvin (gradi centigradi al di sopra dello zero assoluto). La lega è in avvolgimenti raggruppati in bobine e immersi in una matrice di rame. Il conduttore che ne risulta è ulteriormente collocato in una sede in rame che ha la funzione di assicurarne la stabilizzazione e di consentire al magnete di funzionare ininterrottamente anche nell'eventualità che qualche malfunzionamento provochi la perdita di superconduttività. Valori tipici della densità di corrente sono compresi tra 20 000 e 40 000 ampere per centimetro quadrato. Normalmente, tra gli strati di conduttore si interpongono supporti in acciaio inossidabile, e fra i due tipi di strati si alterna uno strato isolante. I supporti in acciaio sono circondati da un dewar (ossia un recipiente di raffreddamento) in acciaio inossidabile.

Il refrigerante è elio liquido, che bolle a 4,2 kelvin. Normalmente si praticano dei passaggi per l'elio, in modo che il conduttore venga a contatto con un bagno di elio. Il calore generato nel conduttore provoca l'ebollizione dell'elio e il vapore risultante asporta il calore prodotto. Il vapore viene quindi condensato in un refrigeratore. Un refrigeratore di elio liquido di concezione moderna richiede circa 500 watt di potenza per la rimozione di un watt di calore emesso a 4,2 kelvin. È peraltro previsto che per un reattore elettrogeneratore da un milione di chilowatt la dissipazione di potenza da parte dei refrigeratori dei magneti sia modesta, dell'ordine di 10 000 chilowatt.

Una parte rilevante dei problemi relativi al progetto dei magneti è di carattere strutturale. Gli avvolgimenti dei magneti superconduttori possono essere considerati alla stregua di grossi recipienti a pressione per il contenimento dell'energia immagazzinata nel campo magnetico. Nei tokamak e negli specchi tandem tale energia ammonta a 50-100 miliardi di joule. Le forze connesse a questa energia sono sia complanari sia normali agli avvolgimenti. Per esempio, in un tokamak la componente toroidale del campo magnetico (ossia la componente concentrica al toro) è prodotta dagli avvolgimenti circolari o a forma di lettere *D*, attraverso i quali passa il plasma toroidale. Il carico planare produce su ciascun avvolgimento una forza radiale di circa 1050 chilogrammi per centimetro quadrato che mette in tensione l'avvolgimento. Contemporaneamente un carico non complanare variabile con la posizione tende a capovolgere l'avvolgimento stesso.

Tutto sta a indicare che lo sforzo finora effettuato per lo sviluppo di magneti superconduttori stabili e stazionari renderà sicuramente tali magneti disponibili per l'impiego nei reattori a fusione. Per contro, i magneti che generano campi variabili al tasso di 10 000 gauss al secondo comportano problemi molto più complessi. I campi variabili generano forze dipendenti dal tempo e inoltre inducono correnti parassite nella struttura degli avvolgimenti. Le correnti parassite dissipano calore nell'elio refrigerante, che funge da termovettore, e aggravano il carico di refrigerazione. Ma i reattori necessiteranno di campi variabili? La risposta è negativa per i reattori a specchio tandem ed è dubitativa per i tokamak, a meno che non si riesca, mediante le onde di radiofrequenza, a pilotare la corrente del plasma fino ai valori di esercizio. È questo un esempio di come i risultati di esperienze fisiche potranno incidere profondamente sulla tecnologia (e persino, forse, sulla fattibilità) dei sistemi di confinamento.

Un secondo aspetto ingegneristico comune a tutti i sistemi di confinamento consiste nella necessità di riscaldamento ausiliario per elevare la temperatura del plasma. Si può anche sfruttare il calore resistivo derivante dalla conduzione di corrente da parte del plasma, che però normalmente risulta insufficiente. Le più comuni tecniche di riscaldamento ausiliario impiegano onde di radiofrequenza o fasci intensi di atomi o molecole neutri di alta energia.

Il riscaldamento a radiofrequenza ha luogo nella conversione in energia termica di onde elettromagnetiche o elettrostatiche di radiofrequenza, provocata da una interazione di risonanza tra le onde e le particelle del plasma. L'attenzione è attualmente puntata in particolare su tre regimi di frequenze. Nel primo, tra 50 e 100 megahertz, la frequenza delle onde corrisponde alla frequenza di ciclotrone degli ioni o a una sua armonica; nel secondo, tra uno e tre gigahertz, alla frequenza delle fluttuazioni di densità degli ioni nel plasma; nel terzo, tra 50 e 100 gigahertz, alla frequenza di ciclotrone degli elettroni.

Tutti i sistemi di riscaldamento a radiofrequenza sono essenzialmente analoghi a stazioni trasmittenti e sono costituiti da una sorgente di radiofrequenza, un amplificatore, una linea di trasmissione e una struttura di emissione posta sulla parete della camera del plasma o internamente a essa. Per la sorgente e l'amplificatore si impiegano tubi elettronici, klystron o maser, a seconda del regime di frequenza; la linea di trasmissione è un cavo coassiale o una guida d'onda e l'emettitore è un'antenna oppure una guida d'onda. L'emettitore deve essere rinforzato contro il danneggiamento da parte dei neutroni o di altre radiazioni emesse dal plasma. Inoltre, occorre interporre una finestra dielettrica nella linea di trasmissione per consentire il passaggio delle onde e contemporaneamente impedire che il trizio, che è radioattivo, sfugga entro il sistema.

Il metodo alternativo di riscaldamento del plasma prevede l'impiego di fasci intensi di particelle neutre (atomi o molecole) di alta energia. Questo metodo è fondamentale, per esempio, per i reattori a specchio tandem. Le particelle neutre attraversano inalterate il campo magnetico del reattore e vengono ionizzate e catturate negli urti con gli ioni e gli elettroni del plasma. Il plasma si riscalda quindi man mano che gli ioni veloci vengono rallentati nella successione di urti. Il fascio stesso origina come plasma prodotto da un generatore a radiofrequenza o da una scarica ad alta intensità di corrente. Normalmente è costituito da atomi ionizzati di idrogeno e di deuterio e in buona parte da una miscela di molecole biatomiche e triatomiche ionizzate. All'accelerazione del plasma in uno o più stadi provvede una serie di griglie metalliche, che devono essere raffreddate con continuità e non devono perturbare il fascio. Lo scopo è quello che le particelle si muovano parallelamente tra loro e assumano la stessa velocità.

Le particelle vengono quindi neutralizzate facendo attraversare al fascio una camera contenente un gas freddo di idrogeno o di deuterio neutri. Nelle cosiddette reazioni di scambio di carica, ciascuno ione cattura un elettrone da un atomo neutro freddo. Gli atomi neutri freddi, così trasformatisi in ioni, assumono un moto di deriva verso una parete della camera, ove vengono raccolti. Nel frattempo il fascio è fatto passare entro un tubo di propagazione e la piccola frazione residua di ioni viene deflessa da un magnete. Il fascio principale, costituito di atomi neutri, attraversa il campo magnetico del reattore e penetra nel plasma.

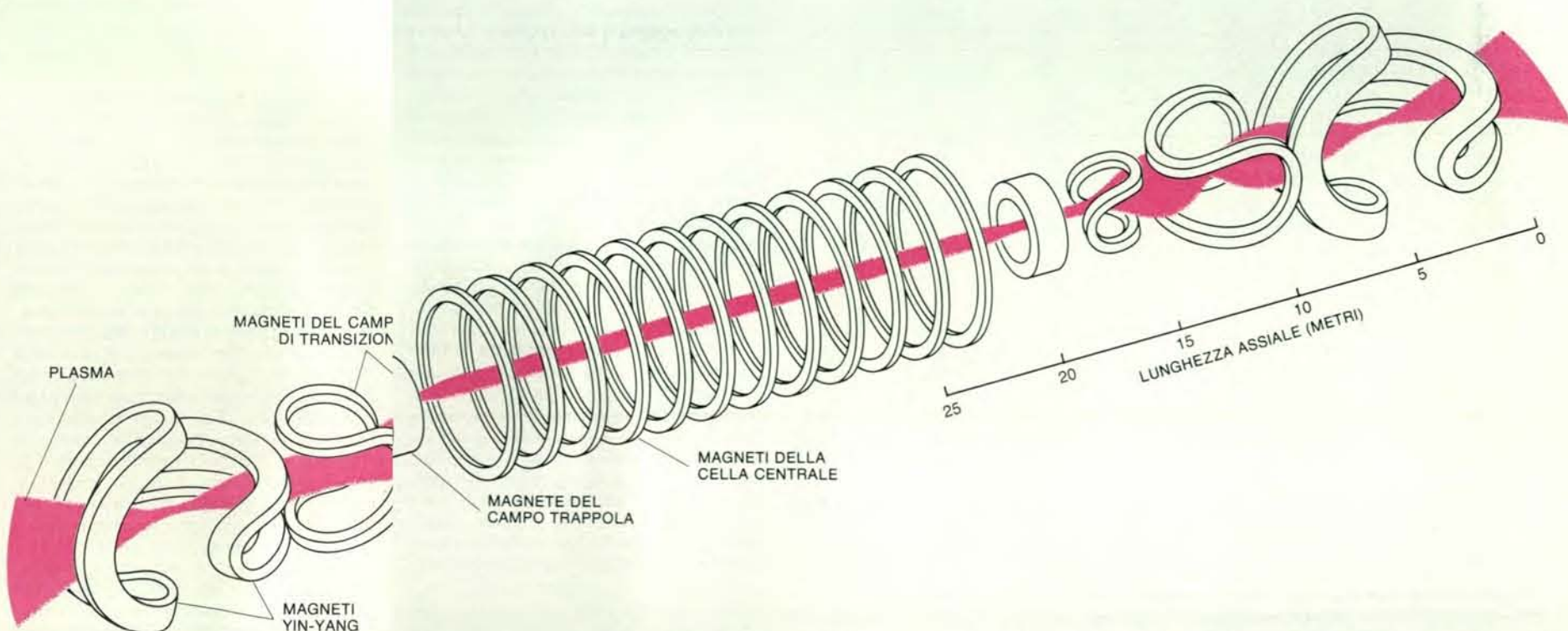
L'insieme delle condizioni di energia, potenza e corrente in cui devono funzionare i reattori a fusione rende complessa la tecnologia dei sistemi di riscaldamento ausiliario mediante fasci neutri. Per un'adeguata penetrazione del fascio entro il plasma di fusione occorre un'energia del fascio compresa tra 100 e 200 chilovolt. Poiché il riscaldamento del plasma può richiedere una potenza fino a 100 megawatt, a un fascio di particelle da 200 chilovolt dovrà corrispondere una corrente di 500 ampere e una densità di corrente di circa un ampere al centimetro quadrato. Queste prestazioni esorbitano nettamente rispetto a quelle dei normali acceleratori di particelle da ricerca, nei quali la corrente a 200 chilovolt è dell'ordine dei millesimi di ampere.

Nondimeno, lo sviluppo registratosi nell'ingegneria dei fasci neutri è molto rilevante. All'inizio degli anni settanta, i sistemi a fasci neutri messi a punto all'Oak Ridge National Laboratory e al Lawrence Berkeley Laboratory fornivano un fascio con un'energia di circa 10 chilovolt e della potenza di circa 100 chilowatt; la durata dell'impulso era brevissima: inferiore a 0,05 secondi. Nelle esperienze attualmente in corso, ai fasci con un'energia di oltre 100 chilovolt è associata una potenza di 10 000 chilowatt, in impulsi di durata prossima al mezzo minuto.

Il calore trasmesso al plasma dal sistema di riscaldamento ausiliario è destinato a fuoriuscire insieme con quello trasmesso al plasma dalle particelle alfa. Quantitativamente, in un reattore della potenza elettrica di 1,2 milioni di chilowatt, si può prevedere una riemissione dell'ordine di 0,5 - 1 milione di chilowatt. È evidente che i problemi maggiori riguardano l'ingegneria dei componenti posti in prossimità del plasma e destinati alla trasmissione del calore.

Le perdite termiche possono assumere aspetti diversi. Per esempio, il plasma emette radiazioni elettromagnetiche (essenzialmente raggi X) e gli atomi delle impurezze presenti nel plasma possono incrementare fortemente l'entità di tali radiazioni. Si presta quindi attenzione a preservare la purezza del plasma e a minimizzare l'erosione della prima parete (la superficie interna della camera del plasma) e delle varie strutture interne alla camera stessa. Nelle configurazioni aperte, quali gli specchi tandem, le perdite di particelle cariche si hanno soprattutto alle estremità del plasma e, quindi, è possibile intervenire su queste particelle, stando a distanza dalla camera del plasma.

Diversa è la situazione per i reattori toroidali, nei quali le perdite maggiori sono rappresentate dalle particelle cariche che diffondono radialmente dal centro alla periferia del plasma. Tali particelle vanno a interagire con un limitatore, ossia un dispositivo in grado di controllare una perdita termica dell'ordine di 10 milioni di watt per metro quadrato. Il limitatore è collocato all'interno della prima parete ed è materialmente in contatto con il plasma. In alcuni sistemi, una particolare configurazione del campo magnetico provoca la diversione delle particelle del plasma, e del calore a queste connesso, dalla camera del plasma verso una «zona di divertore», in cui il calore può essere rimosso da collettori a piastra e il riflusso di gas neutro e di impurezze può essere ridotto al minimo.

Per quanto riguarda il problema dell'erosione della prima parete sotto gli urti delle particelle del plasma, recenti esperienze hanno dimostrato che una soluzione è possibile se per il limitatore si impiegano materiali di basso numero atomico, quale la grafite (carbonio). (In un plasma l'emissione di radiazione da parte delle impurezze tende a crescere con il crescere del numero atomico delle stesse.) È però probabile che, internamente alla prima parete dei reattori, non sia possibile l'impiego della grafite, a causa delle reazioni chimiche tra idrogeno e carbonio. Ai Sandia National Laboratory e in altri centri è allo studio un materiale speciale, del tipo del carburo di silicio e del carburo di titanio, per il rivestimento dei componenti destinati a resistere sia a un flusso termico sia a un bombardamento di ioni del plasma.

Il sistema di confinamento del plasma a specchio tandem è un'alternativa al tokamak. Esso si fonda in parte su campi magnetici e in parte sul campo elettrostatico connesso alla distribuzione di carica elettrica nel plasma. Sono indicati i magneti necessari. La cella centrale è circondata da magneti anulari e alle sue estremità sono presenti un magnete del campo trappola, una coppia di magneti del campo di transizione e una coppia di magneti della configurazione yin-yang. La scala a destra corrisponde alla scala dei grafici della pagina successiva.

Consideriamo ora la produzione di energia da parte di un reattore di potenza. Nella reazione tra un deutone e un tritone solo un quinto dell'energia emessa viene asportato da una particella alfa; gli altri quattro quinti competono a un neutrone da 14 milioni di elettronvolt (14 MeV), che fuoriesce liberamente dal plasma e penetra nella prima parete e nella zona circostante, che costituisce il mantello. Il mantello ha due funzioni: innanzitutto, quella di recuperare l'energia del neutrone. Il percorso tipico di un neutrone da 14 MeV negli urti con i nuclei dei solidi o dei liquidi è intorno ai 10 centimetri. A ogni urto, l'atomo urtato riceve energia dal neutrone, la cui traiettoria viene modificata, con conseguente deposito locale di calore nel mantello. Il termovettore che attraversa il mantello rimuove il calore e lo trasmette a un impianto di potenza relativamente tradizionale, costituito da scambiatori di calore, generatori di vapore e turbine. Il neutrone può inoltre venire assorbito da un nucleo, trasformando l'atomo in un atomo diverso, che può essere stabile o radioattivo. Il flusso neutronico si attenua con la distanza dalla prima parete cosicché il mantello ha uno spessore compreso tra 60 e 80 centimetri. I neutroni che sfuggono dal retro del mantello sono catturati da uno schermo progettato espressamente per proteggere i magneti superconduttori dal danneggiamento da radiazioni e dal riscaldamento eccessivo.

La seconda funzione del mantello è quella di produrre combustibile per il reattore. Il deuterio è abbondantissimo (nell'acqua normale è presente nella percentuale di una parte su 5000 mentre il trizio, essendo radioattivo con una vita media di 12,35 anni, è piuttosto raro in natura. (I raggi cosmici producono piccole quantità di trizio in reazioni che si svolgono nell'alta atmosfera.) Per fortuna, il trizio può essere prodotto artificialmente: in particolare, si impiegano i neutroni emessi dalle reazioni di fusione per provocare reazioni nucleari con i due isotopi naturali del litio: litio 6 e litio 7. Il litio 6 è in grado di catturare neutroni di qualsiasi energia e di produrre trizio ed elio; il litio 7 cattura invece neutroni di energia superiore a 2,87 MeV e produce trizio, elio e un altro neutrone, più lento. Il rapporto di fertilizzazione del reattore (numero di tritoni prodotti per tritone consumato) dipende dalle caratteristiche del progetto, ma può arrivare fino a un valore di 1,5.

Gli ingredienti essenziali del progetto del mantello sono quindi la struttura, il termovettore e il materiale fertile per la produzione del trizio. Per la struttura, i materiali più adatti sono l'acciaio inossidabile, gli acciai ferritici o ad alto tenore di manganese e le leghe di vanadio. Per il termovettore la scelta è tra l'acqua, un gas (quale l'elio) o un metallo liquido. Tra i metalli liquidi i favoriti sono il litio e i composti e gli eutettici di litio e piombo. Il litio liquido è interessante per il basso punto di fusione (186 gradi centigradi), per l'ottima capacità di trasferimento del calore e per la buona compatibilità chimica con la maggior parte delle leghe strutturali; inoltre, il litio funge da materiale fertile per la produzione del trizio.

Il trizio può essere estratto dal flusso di una parte del litio termovettore. Essendo il trizio sotto forma di triziuro di litio, un possibile metodo di estrazione consiste nel porre il flusso di litio in contatto con un eguale volume di sale fuso, costituito per esempio da una miscela di fluoruro, cloruro e bromuro di litio. Il triziuro di litio tende così a mescolarsi al sale, ma ne può essere facilmente separato perché il suo peso è inferiore di tre-quattro volte a quello del sale. Il trizio viene quindi estratto per elettrolisi. Nel suo insieme, tale processo è una versione semplificata del ritrattamento a sali fusi, sviluppato per il combustibile dei reattori a fissione.

Materiali fertili alternativi per la produzione del trizio sono i composti ceramici, tra cui l'ossido e il silicato di litio. L'impiego di questi materiali presenta due vantaggi principali: la quantità di litio necessaria nel reattore può essere modesta e il composto è chimicamente stabile. Per contro, è difficile ottenere un rapporto di fertilizzazione superiore all'unità. Per esempio, nell'alluminato di litio ($LiAlO_2$), gli atomi di alluminio e ossigeno competono con il litio nel diffondere o assorbire i neutroni. È quindi necessario un moltiplicatore neutronico, da collocarsi subito sul retro della prima parete: esso converte un neutrone molto veloce in due neutroni più lenti attraverso una reazione nucleare. I migliori moltiplicatori neutronici sono il berillio e il piombo.

L'intensità del campo e il potenziale elettrostatico alle estremità di un reattore a specchio tandem sono il risultato dell'azione congiunta dei magneti, del riscaldamento del plasma mediante fasci di particelle elettricamente neutre di alta energia e dell'estrazione mediante pompaggio degli ioni da alcune zone del plasma. Nei diagrammi sono rappresentati l'intensità del campo magnetico e il potenziale elettrostatico alle due estremità di uno specchio tandem. Gli atomi ionizzati del plasma vengono confinati nella cella centrale dal picco magnetico in corrispondenza del magnete del campo trappola. In pratica, sono riflessi da una strozzatura del campo magnetico, che agisce da specchio. Gli ioni che riescono ad attraversare lo specchio rimbalzano allontanandosi dal picco elettrostatico presente nella configurazione yin-yang. Gli elettroni del plasma sono respinti man mano che si approssimano ai due lati della valle elettrostatica entro la zona di transizione. È così possibile procedere all'efficace riscaldamento degli elettroni elettrostaticamente isolati nella cella terminale.

L'iniezione di fasci neutri è una delle tecniche di riscaldamento di un plasma di fusione. Le particelle del fascio (spesso atomi di deuterio), ionizzate e accelerate, catturano elettroni da atomi di deuterio freddi, neutralizzandosi nuovamente. Attraversano così un campo magnetico, penetrano nel plasma e trasferiscono energia per urto. Un'alternativa è il riscaldamento del plasma con radiazioni elettromagnetiche di radiofrequenza.

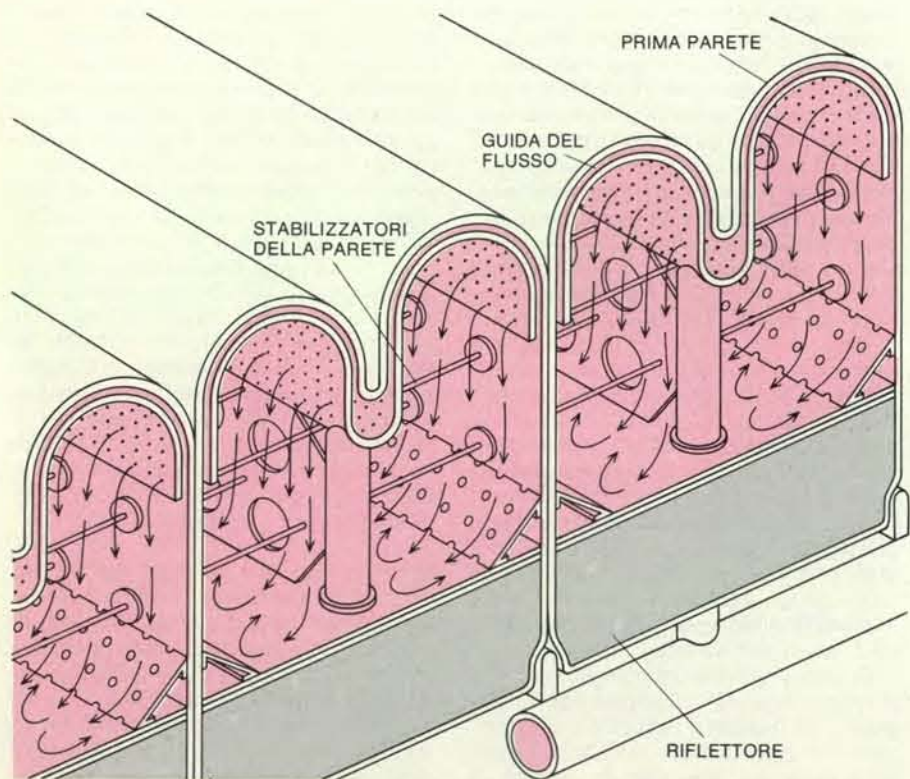
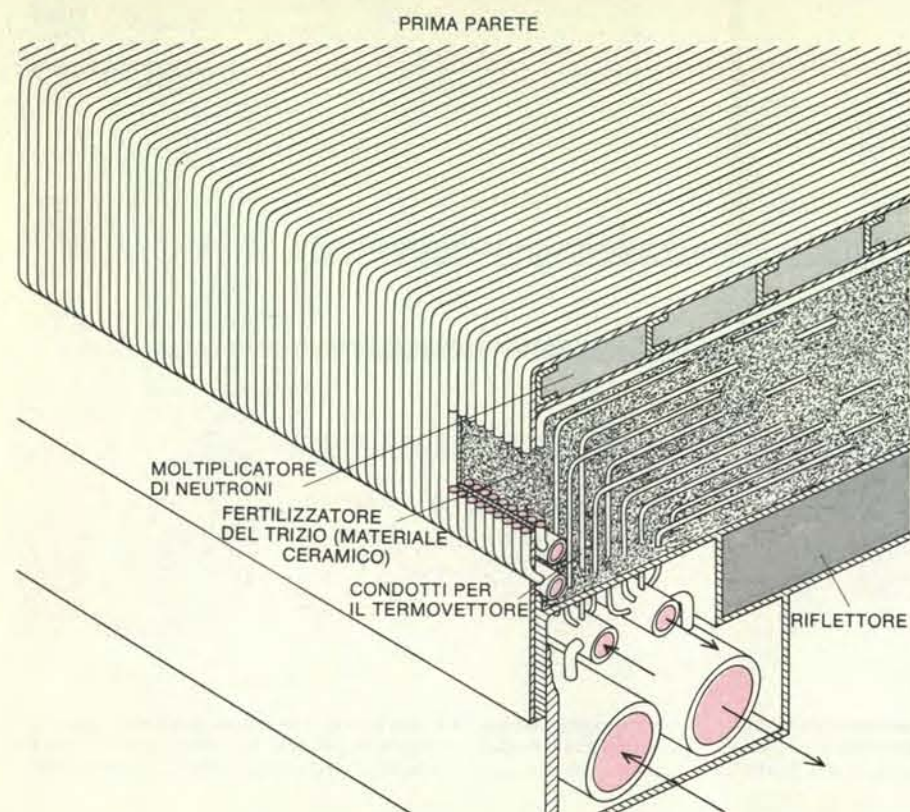
L'estrazione del trizio da un materiale ceramico è complicata perché si fonda sul fenomeno della diffusione da un solido. Normalmente la velocità di diffusione è bassa e quindi occorre realizzare il materiale ceramico in pastiglie i cui granuli abbiano una dimensione dell'ordine dei 10^{-3} millimetri. Inoltre il materiale non deve subire il processo di sinterizzazione neppure approssimandosi o addirittura superando una temperatura di 1000 gradi centigradi. Il trizio che diffonde dal materiale ceramico può venire quindi rimosso dal mantello mediante un flusso di elio gassoso che attraversa il letto di pastiglie. Solo in base all'esperienza pratica sarà però possibile sapere se il sistema di fertilizzazione opererà come da progetto.

Vari componenti del reattore presentano alcuni problemi di ingegneria. Uno di questi è il danneggiamento da irraggiamento neutronico. Quando un neutrone è diffuso o assorbito in un urto, una parte della sua energia è trasmessa all'atomo urtato, il quale rincula provocando lo spostamento, talvolta definitivo, di altri atomi. Questi ultimi, detti atomi interstiziali, si lasciano dietro delle vacanze. Sia le vacanze, sia gli atomi interstiziali, possono aggregarsi in anelli. Se poi è presente un agente stabilizzante quale un gas, le vacanze possono diventare permanenti. Ed è proprio un gas che viene prodotto nei componenti metallici di un reattore a fusione. Per esempio, si produce idrogeno ed elio se neutroni di energia superiore ai 5 MeV interagiscono con gli atomi del metallo. Durante l'interazione alcuni di questi atomi del metallo possono trasformarsi in isotopi radioattivi dello stesso o di altri elementi. In seguito a tutte queste trasmutazioni le proprietà del materiale si modificano. Normalmente il materiale si indurisce e perde duttilità.

Il flusso di neutroni da 14 MeV attraverso la prima parete si misura in megawatt per metro quadrato. Un flusso di 4×10^{17} neutroni per metro quadrato per secondo rappresenta un flusso di potenza, verso la prima parete, di un megawatt per metro quadrato. In pratica, tale flusso di potenza può risultare da tre a cinque volte più elevato. È possibile calcolare il numero di spostamenti all'anno, subiti da un atomo nel reticolo di un materiale, per ogni megawatt per metro quadrato di flusso di potenza. È un numero variabile, che per l'acciaio inossidabile della prima parete è compreso tra 10 e 20. Poiché in pratica tutti gli atomi spostati si assestano definitivamente in qualche punto vacante, l'acciaio non si trasforma in materiale gelatinoso. È comunque chiaro qual è il problema. Lo stesso flusso di potenza dà inoltre luogo a circa 150 atomi di elio e a 500 atomi di idrogeno all'anno per milione di atomi del metallo.

Dopo il problema del confinamento del plasma, l'ostacolo principale che l'ingegneria dei reattori a fusione deve affrontare è forse quello delle modificazioni che avvengono nei materiali in seguito al danneggiamento da radiazioni. Le ricerche in questo settore hanno preso avvio con l'irraggiamento di campioni metallici, effettuato in reattori a fissione quali lo High Flux Isotope Reactor di Oak Ridge, che è in grado di produrre un gran numero di neutroni aventi un ampio spettro energetico. Se il campione è una lega contenente nichel, si producono migliaia di atomi di elio per milione di atomi di metallo e si osservano modifiche nel materiale, quali rigonfiamento e infragilimento. Secondo una valutazione preliminare, in un reattore a fusione in esercizio, commerciale, la prima parete e il mantello mantengono proprietà adeguate per un periodo di tre-sei anni. È quindi necessario che il progetto dei reattori a fusione preveda la sostituzione periodica della prima parete e del mantello nel corso della vita dell'impianto. È altrettanto evidente la necessità di un ulteriore, notevole impegno di ricerca. Attualmente, gli Stati Uniti, con l'eventuale collaborazione di paesi dell'Europa occidentale e del Giappone, stanno prospettando la realizzazione del Fusion Materials Irradiation Test, un apparato sperimentale per la produzione di neutroni da 14 MeV. Tale apparato rappresenterà una parte importante nel programma di sviluppo di materiali adatti per i reattori a fusione.

Sia nei tokamak, sia negli specchi tandem, la perdita di combustibile termonucleare dal plasma è da 10 a 20 volte superiore allo stesso consumo di combustibile. Ne consegue che è essenziale recuperare e riciclare il combustibile. Il recupero di combustibile è particolarmente impegnativo, in quanto il pompaggio deve essere molto elevato, da uno a dieci milioni di litri al secondo, e il vuoto nella camera del reattore deve essere estremamente spinto, intorno a 10^{-7} atmosfere. Tra le alternative tecniche disponibili, la migliore attualmente sembra quella delle pompe a crioassorbimento operanti alla temperatura dell'elio liquido.

In queste pompe è presente una serie di filtri (o stacci) molecolari raffreddati a 4 kelvin dall'elio liquido. Attorno ai filtri è posto un dispositivo a zig zag, raffreddato

I progetti della prima parete e del mantello del reattore variano a seconda del tipo di termovettore e di materiale fertile per la produzione del trizio. Nel disegno in alto, il termovettore è acqua (*in colore*) e il materiale fertile è alluminato di litio o ossido di litio, entrambi materiali ceramici solidi. I neutroni emessi dal plasma di fusione trasformano il litio del materiale ceramico in trizio, elio e neutroni addizionali. Il trizio esce per diffusione dal materiale ceramico e viene rimosso mediante un flusso di elio gassoso. La produzione di trizio viene incrementata collocando un riflettore neutronico (grafite) al di sotto del mantello e un moltiplicatore neutronico (berillio o piombo) al di sopra. Nelle reazioni nucleari che hanno luogo nel moltiplicatore, vengono assorbiti i neutroni veloci provenienti dal plasma ed emessi neutroni più lenti in numero doppio. Nel disegno in basso, il termovettore e il materiale fertile sono entrambi litio liquido (*in colore*). Il trizio viene rimosso per estrazione dal termovettore circolante verso gli scambiatori di calore del reattore. I due disegni della illustrazione derivano dallo studio per il progetto di reattore tokamak STARFIRE.

a 77 kelvin dall'azoto liquido. Sui filtri condensano i gas volatili quali l'idrogeno, il deuterio e il trizio, mentre i gas meno volatili, quali l'ossigeno, l'azoto e il metano, vengono adsorbiti dal dispositivo a zig zag. Si può anche catturare l'elio spruzzando argo liquido sulle superfici delle pompe: l'argo sequestra l'elio rivestendolo. Le pompe a crioassorbimento presentano una alta velocità di pompaggio, ma richiedono la rimozione periodica dei gas che catturano. Inoltre, la cattura produce nel reattore un elevato accumulo di trizio. Per fare fronte a questi inconvenienti sono stati progettati dei «divertori» magnetici o limitatori di pompaggio, che consentono di elevare a 10^{-5} atmosfere la pressione del gas all'imbocco dei condotti di pompaggio. Si possono così utilizzare pompe che non catturano combustibile.

Quando il flusso di scarico è ripulito dalle impurezze (incluse le «ceneri» di elio prodottesi nelle reazioni di fusione), è necessario separare il deuterio e il trizio incombusti in modo che si possano combinare, nelle rispettive proporzioni, con il trizio estratto dal mantello. Molto semplicemente la separazione si effettua per distillazione a temperature criogeniche, sfruttando le differenze tra i punti di ebollizione di H_2, DH, D_2, HT, DT e T_2. Il gas meno volatile è H_2, con un punto di ebollizione a 20,39 kelvin; il più volatile è HT_2, che bolle a 25,04 kelvin. Si ottengono purezze accettabili usando colonne di distillazione con circa 60 stadi.

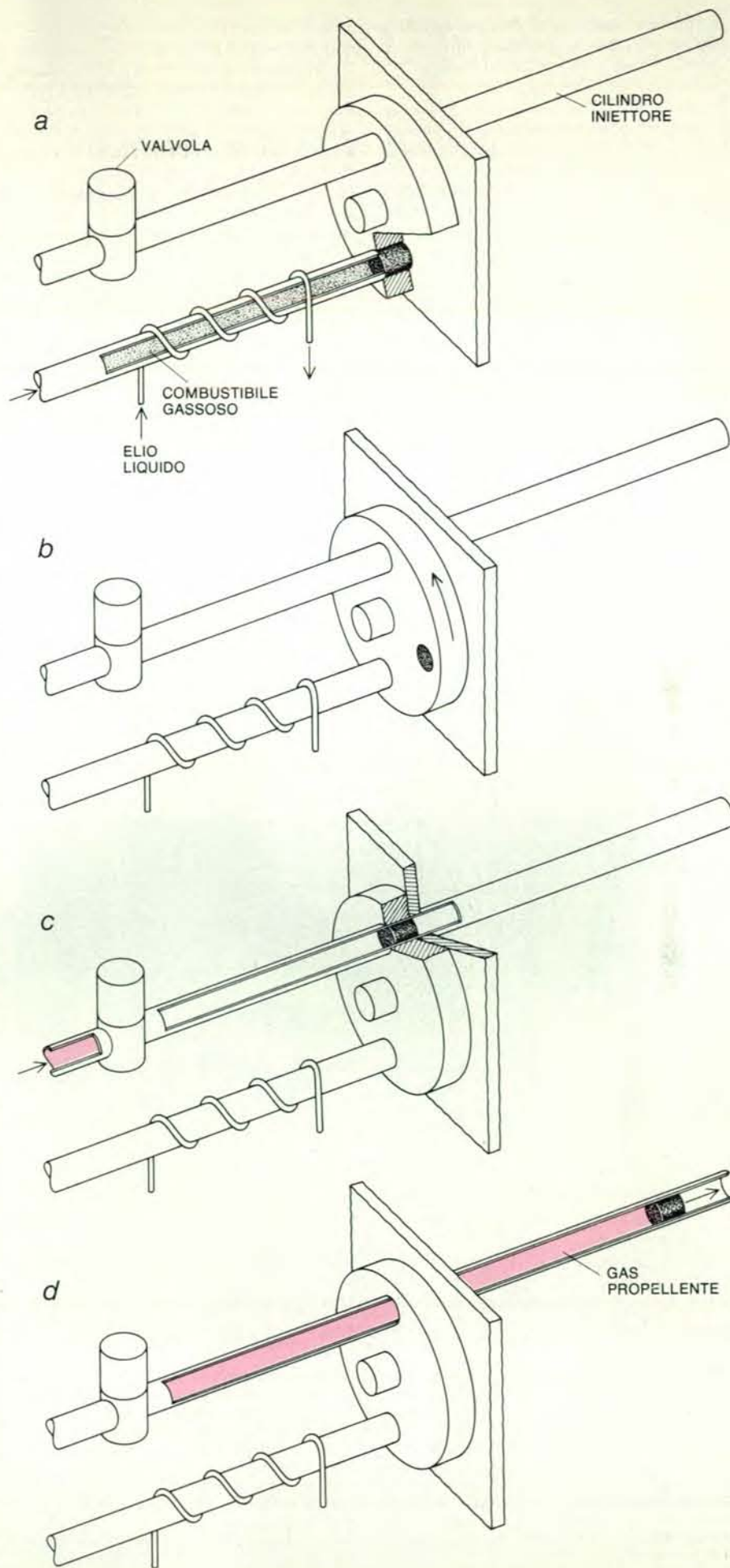
Occorre quindi reiniettare nel plasma la miscela ricostituita di deuterio e trizio. Nelle esperienze in corso con plasmi di idrogeno, si dirigono getti di gas verso la periferia del plasma, che li assorbe agevolmente. Questo assorbimento può non essere altrettanto efficace in un reattore a fusione: in esso gli atomi di gas verrebbero ionizzati entro il plasma fino a non oltre 10 centimetri di distanza dalla periferia del plasma stesso. Occorrerebbero quindi almeno dieci iniettori di gas, costituiti da un ugello e da una valvola rapida, in grado di produrre getti di una durata compresa tra 50 e 100 millisecondi.

In alternativa, si considera l'iniezione ad alta velocità di pastiglie congelate di combustibile, secondo una tecnica la cui fattibilità è stata dimostrata in un tokamak sperimentale a Oak Ridge e che offre ottime garanzie anche per un reattore commerciale. Ciascuna di queste pastiglie di combustibile è assimilabile a una pallina di neve nel fuoco dell'inferno: alla temperatura inferiore al punto di congelamento del deuterio (18,73 kelvin) e del trizio (20,6 kelvin), viene iniettata in un plasma la cui temperatura è di decine di milioni di gradi. Le pastiglie devono avere un diametro di 2-4 millimetri e contenere dal 10 al 15 per cento degli atomi di combustibile complessivamente presenti in ogni istante nel plasma. Le pastiglie vengono accelerate a una velocità di circa due chilometri al secondo e quindi iniettate con una frequenza di 10-20 al secondo. Se si esclude la frequenza di iniezione, gli altri parametri operativi citati sono già stati raggiunti sperimentalmente.

Sono stati proposti due tipi di iniettori. Quelli centrifughi sono costituiti da una piastra rotante ad alta velocità sulla quale sono fissati due tubi a *U*. Una linea di alimentazione raffreddata da elio liquido estrude il combustibile solido entro una delle aperture di ciascun tubo a *U*; quindi il combustibile viene affettato e si produce così una pastiglia all'interno del tubo. La rotazione della piastra accelera la pastiglia, che fuoriesce ad alta velocità dal tubo. Nelle esperienze svoltesi all'Istituto Max Planck per la fisica dei plasmi di Garching e a Oak Ridge, pastiglie di deuterio di un millimetro di diametro sono state accelerate fino a 290 metri al secondo. Iniettori di questo genere dovranno riuscire ad accelerare le pastiglie fino a una velocità di due-cinque chilometri al secondo. Caratteristica intrinseca di questa tecnica è l'alta frequenza di ripetizione.

Sono inoltre in corso di sviluppo gli iniettori a gas: questi dispositivi hanno già consentito di accelerare pastiglie da un millimetro fino alla velocità di un chilometro al secondo. Un progetto costruttivo messo a punto a Oak Ridge prevede un disco con fori da un millimetro, raffreddato con elio liquido e alloggiato in una sede di rame. Il combustibile gassoso congela in ogni foro e il disco viene fatto ruotare in modo che le risultanti barrette di combustibile giungano in corrispondenza di una valvola rapida posta all'imbocco di un cilindro di iniezione lungo 16 centimetri. La valvola libera elio gassoso compresso a 30 atmosfere, che funge da propellente del combustibile. È possibile migliorare le prestazioni del sistema usando idrogeno come propellente e aumentando la lunghezza del cilindro.

In tutte le fasi del ciclo del combustibile è indispensabile garantire il contenimento sicuro nel reattore del trizio, a causa della sua radioattività. Il livello di contenimento accettabile si ottiene con un semplice calcolo. Il previsto accumulo di trizio in un reattore di potenza è dell'ordine di 10 chilogrammi, corrispondenti a circa 100 milioni di curie di radioattività. (Un curie equivale a 30 miliardi di disintegrazioni nucleari al secondo.) Non è stato ancora definito ufficialmente un limite accettabile per il tasso di perdita di trizio, ma gli addetti ai lavori fanno generalmente riferimento a un valore compreso tra uno e dieci curie al giorno. La perdita

L'iniezione di combustibile si può effettuare propellendo nel plasma pastiglie congelate di deuterio e di trizio. Secondo un progetto di iniettore elaborato all'Oak Ridge National Laboratory, un flusso di combustibile gassoso viene refrigerato con elio liquido in modo che una barretta di combustibile, divenuto solido, venga a collocarsi entro un apposito foro praticato in un disco (*a*). Ruotando (*b*), il disco porta la barretta in corrispondenza di una valvola all'imbocco di un cilindro di iniezione (*c*). La valvola libera getti di propellente (*in colore*), che è elio gassoso ad alta pressione (*d*). Per un reattore a fusione, l'iniettore deve essere in grado di impartire alle pastiglie di combustibile una velocità pari a due chilometri al secondo.

giornaliera massima di trizio consentita per un reattore in esercizio normale è quindi di una parte su cento milioni.

Il contenimento del trizio si può ottenere ricorrendo alla presenza di almeno tre barriere in ogni punto dell'impianto in cui si trova il trizio. La prima barriera è l'elemento strutturale di contenimento fisico del trizio, per esempio una tubazione. Viene quindi una barriera di contenimento di tale elemento strutturale: per esempio una seconda tubazione oppure, per un insieme di dispositivi per la manipolazione del trizio, una camera a guanti. Il terzo livello di contenimento è rappresentato dalla sala del reattore, dai locali contenenti le attrezzature per il trattamento del trizio, eccetera. Ciascuno di questi locali deve essere equipaggiato con un sistema di decontaminazione del trizio. Si valuta che, in meno di una giornata, sia possibile rimuovere dall'edificio reattore un chilogrammo di trizio disperso dall'impianto. Al Los Alamos Scientific Laboratory è stata realizzata la Tritium System Test Assembly, un impianto sperimentale per la simulazione dell'intero ciclo del trizio, la cui entrata in funzione è imminente. Le informazioni che deriveranno dalle esperienze previste su questo impianto dovrebbero consentire il trattamento sicuro del trizio nelle prossime apparecchiature sperimentali, e quindi nei reattori.

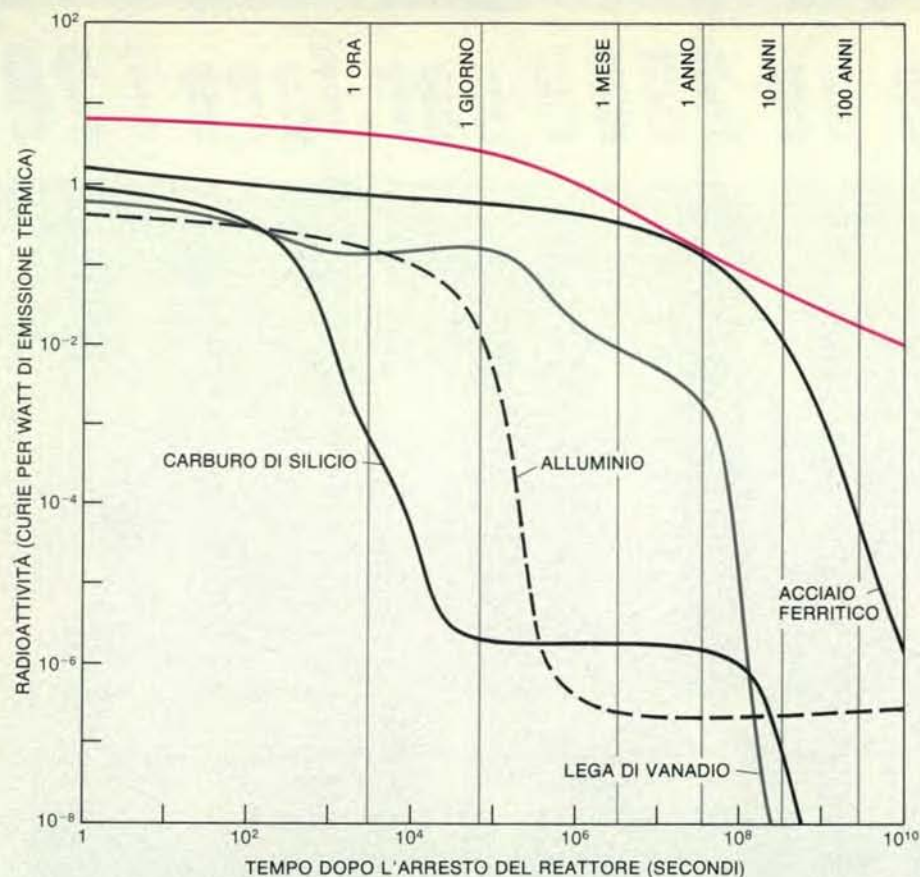
In un reattore a fusione è presente una sola altra sorgente di radioattività: il reattore stesso, e in particolare la prima parete, il mantello e lo schermo, che si attivano in seguito all'irraggiamento da parte del flusso neutronico. L'entità e la natura della radioattività dipendono dai materiali costitutivi del reattore. Un'utile valore di merito è il potenziale di rischio biologico o BHP (da *Biological Hazard Potential*). La concentrazione massima ammissibile per i vari radionuclidi nell'aria e nell'acqua è specificata da un'apposita norma federale e il BHP rappresenta il rapporto tra il valore misurato di radioattività (espresso in curie per unità di volume) e il livello federale consentito. Per tutti i materiali di interesse pratico il BHP di un reattore a fusione risulta di almeno un fattore dieci, e in molti casi di oltre un fattore cento, inferiore al BHP di un reattore a fissione.

È però forse più importante sapere se i materiali di rifiuto di un reattore a fusione si possano eliminare definitivamente mediante interramento a bassa profondità. Per tale tipo di sistemazione sono state recentemente proposte alcune direttive federali, in base alle quali i rifiuti sono ripartiti in tre classi. Alla classe *A* appartengono i rifiuti che raggiungono un livello di radioattività di sicurezza accettabile dopo non più di dieci anni di interramento a bassa profondità. In termini quantitativi, la situazione di «sicurezza accettabile» è definita come quella di una persona che, accedendo inavvertitamente al sito di interramento dei rifiuti e permanendovi, riceva una dose annuale di radiazione non superiore a cinque volte il livello del fondo naturale di radiazione. Alla classe *B* appartengono i materiali chimicamente stabilizzati, che pervengono a un livello di «sicurezza accettabile» entro 100 anni. Tali materiali devono mantenere la propria stabilità e venire interrati in modo che la dose assorbibile da parte di una persona presente sul sito non superi che di qualche unità per cento il livello naturale. Alla classe *C* appartengono invece materiali chimicamente stabilizzati che pervengono al livello di sicurezza accettabile entro 500 anni. Tali materiali devono essere interrati ad almeno cinque metri di profondità e il sito deve essere protetto da barriere naturali o artificiali. Per i materiali che non rientrano in queste categorie occorrerà presumibilmente ricorrere a metodi di sistemazione più complessi, quali il deposito in profondità in adatte formazioni geologiche.

I reattori a fusione possono essere realizzati con materiali destinati a interramento a bassa profondità. In particolare, la teoria dimostra che i reattori a fusione in acciaio al ferro-cromo-manganese daranno origine a materiali di rifiuto della classe *C*. Se poi il materiale costruttivo è dato da leghe di titanio, vanadio o alluminio, o da prodotti ceramici, come la grafite e il carburo di silicio, i rifiuti potrebbero essere di classe *B*. Inoltre, non è escluso che si riesca a trasformare leghe aventi buone proprietà meccaniche, ma con tendenza a un'elevata attivazione, in altre che soddisfino almeno i requisiti di appartenenza alla classe *C*. Con tale obiettivo, è in atto un'attività sperimentale volta all'eliminazione del nichel e del molibdeno dalle leghe di acciaio.

Nel lungo termine si potrebbe addirittura riuscire a regolare la composizione isotopica di un materiale in modo da modificarne completamente la tendenza all'attivazione, pur lasciandone inalterate le proprietà meccaniche e chimiche. Si consideri ancora il molibdeno, che nella maggior parte degli acciai impartisce resistenza alle alte temperature e che normalmente costituisce l'uno per cento delle leghe. Gli isotopi naturali del molibdeno, che reagiscono con i neutroni formando elementi radioattivi di lunga vita in un reattore a fusione, sono quelli di peso atomico 92, 94, 98 e 100. Se, per ipotesi, il molibdeno contenuto nell'acciaio del reattore fosse costituito solo dai tre restanti isotopi naturali, di peso atomico 95, 96 e 97, la sua presenza nell'acciaio non pregiudicherebbe l'appartenenza alla classe *C* dei corrispondenti materiali di rifiuto.

Dal punto di vista economico restano ancora delle incognite sulla separazione isotopica; le tecniche attuali sono state infatti sviluppate unicamente per la separazione degli isotopi dell'uranio. Per la separazione di elementi più leggeri, quali nichel e molibdeno, si potrebbe impiegare una nuova tecnica, che viene sviluppata dalla TRW, Inc. Per l'uranio arricchito al

Il danneggiamento da irraggiamento neutronico a cui sono sottoposti i materiali dei reattori a fusione è stato simulato collocando un campione di acciaio inossidabile nel nocciolo di un reattore a fissione da ricerca, alla temperatura di circa 590 gradi centigradi, e mantenendolo per un anno esposto a un elevato flusso neutronico. Il flusso integrato ($1,9 \times 10^{26}$ neutroni per metro quadrato) ha provocato in media, per ogni atomo del reticolo dell'acciaio, dieci spostamenti. Alcune delle vacanze venutesi così a creare nel reticolo si sono poi aggregate in zone vuote (*aree bianche*); gli atomi spostati si sono in parte aggregati in zone denominate anelli di dislocazione (*aree in grigio di media intensità*). Inoltre, gli elementi chimici dell'acciaio si sono separati (*aree in grigio intenso*). Di conseguenza, l'acciaio si è rigonfiato ed è diventato meno duttile. La microfotografia elettronica è stata ottenuta a Oak Ridge da F. W. Wiffen, J. O. Stiegler ed E. E. Bloom, con un ingrandimento di 110 000 diametri.

La diminuzione della radioattività indotta dai neutroni nei materiali di un reattore a fusione varia da un materiale all'altro (*curve un nero*); ciò indica che si potrebbe regolare la composizione isotopica di materiali quali le leghe in modo da ridurne la pericolosità pur mantenendone inalterate le proprietà strutturali. In generale, la diminuzione di radioattività in un reattore a fusione è molto più rapida rispetto a quella di un reattore a fissione, come dimostra il confronto, riportato in questo grafico, in cui la curva in colore si riferisce alla diminuzione di radioattività del materiale di rifiuto di un reattore autofertilizzante veloce a metallo liquido.

3 per cento nell'isotopo U-235 è stato definito un obiettivo di costo di 220 dollari al chilogrammo. Anche se la regolazione isotopica del molibdeno portasse a un costo quattro volte superiore, il costo del mantello di un reattore a fusione in acciaio al molibdeno isotopicamente regolato non supererebbe del 5 per cento il costo del mantello in acciaio al molibdeno normale.

Di recente sono stati messi a punto due progetti di reattori a fusione, uno di tipo tokamak, l'altro di tipo a specchio tandem. Entrambi danno un'idea dei possibili reattori del futuro. Il progetto del tokamak, denominato STARFIRE, è stato sviluppato dall'Argonne National Laboratory, in collaborazione con la McDonnell-Douglas Astronautics Company, la GA Technologies e la Ralph M. Parsons Company. Durato due anni, con un costo di 2 milioni di dollari, riguarda un impianto da 1200 megawatt di imponenti dimensioni: il plasma toroidale ha un diametro maggiore di 14 metri e una sezione trasversale di 4,2 metri. Al confinamento del plasma provvedono i magneti superconduttori. Dodici magneti di questo tipo a forma di anello, attraversati dalla colonna di plasma, generano il campo toroidale dell'intensità massima di 111 000 gauss e complessivamente immagazzinano un'energia di 50 miliardi di joule.

Nello STARFIRE il plasma è alimentato da una combinazione di getti gassosi e di pastiglie, è riscaldato mediante onde di radiofrequenza e conduce una corrente di 10 milioni di ampere. La corrente viene sostenuta mediante onde di radiofrequenza, secondo un processo la cui fattibilità non è però ancora pienamente dimostrata. In assenza di sostentamento esterno, sarebbe necessario interrompere la combustione del plasma periodicamente, probabilmente ogni venti minuti o anche con maggiore frequenza, per consentire il ripristino dei magneti, che rappresentano il primario di un trasformatore. Nel corso di tale operazione, avrebbe luogo un raffreddamento della camera del reattore e del mantello, con conseguente modifica della distribuzione degli sforzi nei materiali. La fatica meccanica risultante contribuirebbe a limitare nel tempo l'integrità delle strutture. Il pilotaggio a radiofrequenza della corrente ha però un costo, che è il consumo di energia. Si valuta che occorrano 150 megawatt, cioè oltre il dieci per cento dell'energia elettrica prodotta dall'impianto, per l'alimentazione del sistema a radiofrequenza. Il rendimento netto dei futuri reattori tokamak dovrà migliorare rispetto a quello dell'attuale STARFIRE (30 per cento).

Nello STARFIRE, il mantello è raffreddato con acqua pressurizzata a 155 chilogrammi per centimetro quadrato. L'acqua entra nel reattore a una temperatura di 280 gradi centigradi, assorbe calore circolando entro il mantello e fuoriesce a 320 gradi. Attraversando il mantello passa in tubazioni di acciaio inossidabile immerse nel materiale ceramico fertilizzatore del trizio. Uno strato di piombo-zirconio sulla superficie interna del mantello funge da moltiplicatore neutronico. Ciononostante, si valuta che il rapporto di fertilizzazione del reattore non superi l'unità per più del 5 per cento, il che è probabilmente insufficiente. Vengono pertanto sviluppati altri progetti alternativi.

Il progetto concettuale di reattore a specchio tandem, il Mirror Advanced Reactor Study (MARS), viene sviluppato dal Lawrence Livermore National Laboratory, in collaborazione con la TRW, la General Dynamics Corporation e l'Università del Wisconsin. Dopo due anni di lavoro, il progetto è prossimo al completamento. Esso prevede un lungo reattore lineare, con una cella centrale del plasma di 129 metri di lunghezza e due celle terminali aventi funzione di specchio, che insieme aggiungono altri 48 metri. La cella centrale contiene 75 metri cubi di plasma e genera una potenza continua di 2575 megawatt. La funzione di specchio delle celle terminali è sostenuta mediante iniezione continua di fasci neutri e di calore a radiofrequenza. La cella centrale è alimentata con getti di gas e con pastiglie di combustibile. Un plasma che costituisce un alone attorno al plasma cilindrico principale impedisce la penetrazione di impurezze.

Nella cella centrale del MARS, sono presenti 41 magneti anulari superconduttori, che generano un campo di 47 000 gauss. Altri magneti superconduttori delle celle terminali, del tipo yin-yang, generano un campo di intensità massima di 75 000 gauss; infine, un magnete speciale, costituito da magneti non superconduttori raffreddati ad acqua e da un magnete superconduttore, produce un «campo trappola» di 240 000 gauss tra la cella centrale e ciascuna cella terminale. Attualmente un campo di tale intensità non può ancora essere ottenuto su grandi volumi soltanto con magneti superconduttori. Il campo magnetico complessivo contiene un'energia di circa 50 miliardi di joule. Occorre energia per l'azionamento dei magneti non superconduttori e per refrigerare i superconduttori, per la circolazione del termovettore, per il sostentamento del plasma delle celle terminali e per il funzionamento delle attrezzature ausiliarie. In pratica, il 20 per cento dell'energia prodotta dal reattore deve essere utilizzata nell'impianto. Ciononostante, il rendimento netto del reattore è del 35 per cento.

A un tale valore di rendimento contribuiscono tre caratteristiche del progetto: la presenza della barriera elettrostatica di separazione degli elettroni ad alta tempe-

ratura nelle celle terminali da quelli a temperatura inferiore nella cella centrale; l'impiego di convertitori diretti per la produzione di energia elettrica dall'energia delle particelle cariche perse dalle estremità del reattore; l'impiego di un termovettore ad alta temperatura, il piombo-litio, che entra nel mantello a 350 gradi centigradi e fuoriesce a 500 gradi. Non è però ancora stata ottenuta la conferma sperimentale della praticabilità di queste soluzioni progettuali nel loro complesso. Inoltre, la fisica degli specchi tandem si trova a uno stadio più arretrato rispetto alla fisica dei tokamak. Ne consegue che il MARS soddisfa un maggior numero di condizioni rispetto allo STARFIRE.

Una di queste condizioni riguarda la barriera elettrostatica di separazione degli elettroni. Tale barriera costituisce una zona di basso potenziale elettrostatico, che respinge gli elettroni su entrambi i lati. Verso tale zona sono attratti gli ioni positivi: se questi non vengono rimossi, la barriera si annulla. Nel progetto del MARS è stata quindi prevista una tecnica di rimozione denominata pompaggio dell'orbita di deriva. Per ora si tratta di una tecnica puramente teorica, che consiste in una modulazione dello 0,03 per cento dell'intensità del campo magnetico delle celle terminali mediante bobine semilunari, alla frequenza di circa 100 chilohertz. Tale modulazione provoca un moto radiale di deriva degli ioni diretto verso l'esterno; una volta usciti dalla portata della trappola magnetica ed elettrostatica, gli ioni si perdono all'estremità della cella.

I progetti dei tokamak e degli specchi tandem differiscono per vari aspetti, connessi talvolta alla necessità di adottare tecniche particolari, quali la conversione diretta. Numerose sono però le caratteristiche comuni: per esempio le tecniche del vuoto, la gestione del combustibile, i magneti superconduttori, il riscaldamento del plasma e l'iniezione del combustibile. Sia lo STARFIRE che il MARS forniscono indicazioni a favore della realizzazione di reattori a fusione ed entrambi mostrano chiaramente dove siano necessari miglioramenti progettuali.

Alla domanda se sia fattibile la fusione a confinamento magnetico non si è ancora in grado di dare una risposta definitiva, anche se tale risposta non tarderà a venire, in considerazione dello stato attuale delle ricerche. Gli Stati Uniti, i paesi dell'Europa occidentale, il Giappone e l'Unione Sovietica hanno tutti in corso programmi specifici di sviluppo ingegneristico e hanno preso parte a uno studio multinazionale per far progredire la progettazione dei reattori tokamak (INTOR). Un accordo bilaterale o multilaterale potrebbe forse agevolare i prossimi importanti sviluppi sperimentali. In caso contrario, occorrerà comunque potenziare i programmi sperimentali attualmente in corso. In entrambi i casi, si assisterà nel prossimo decennio a un notevole ampliamento delle conoscenze sull'ingegneria dei reattori a confinamento magnetico.

La struttura su grande scala dell'universo

Un alveare di superammassi galattici e ampi vuoti si stende nello spazio per miliardi di anni luce: questa struttura può derivare da perturbazioni nella densità della materia sviluppatesi durante le prime fasi del big bang

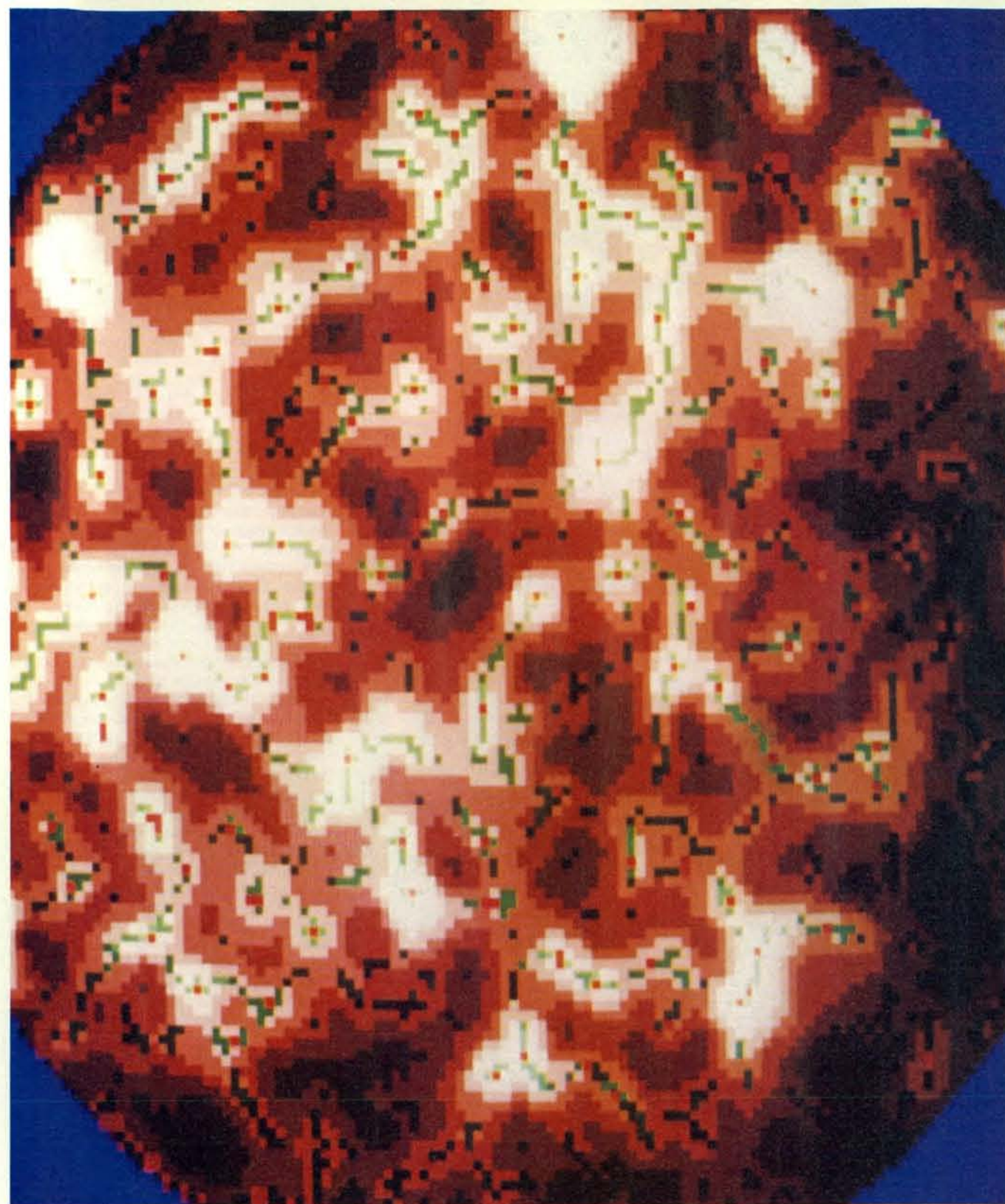
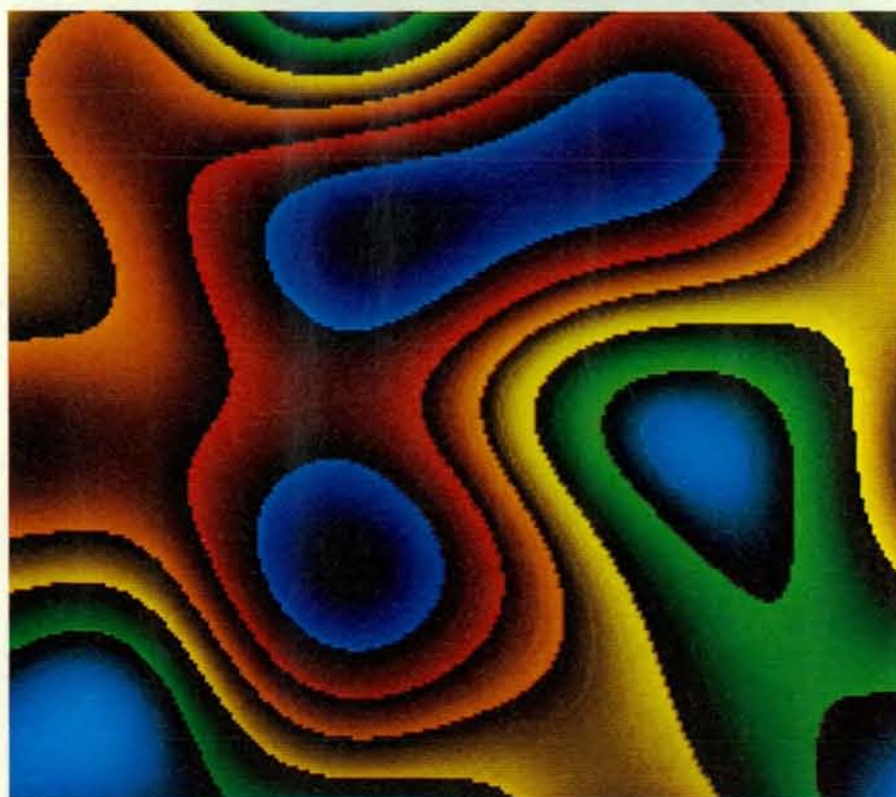
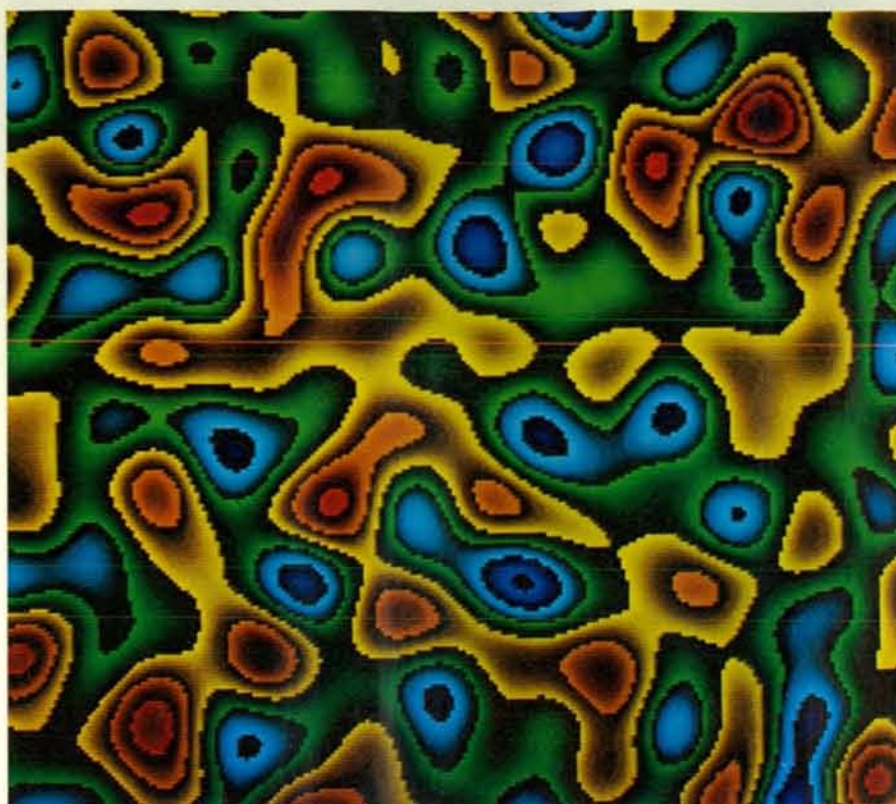
di Joseph Silk, Alexander S. Szalay e Yakov B. Zel'dovich

Gli astronomi si sono resi conto da molto tempo del fatto che la distribuzione della materia su scala cosmica deve in qualche modo portare il segno di una fase molto antica nella storia dell'universo. Di questa distribuzione e del suo evolversi è necessario elaborare una spiegazione coerente nel contesto della teoria del big bang, dal momento che tra i cosmologi e gli astrofisici esiste un consenso pressoché universale sul fatto che il big bang fornisca una cornice empirica complessiva all'interno della quale è possibile esaminare tutte le questioni cosmologiche. Secondo la teoria del big bang l'universo ha avuto inizio come punto singolare di densità infinita circa 10 o 20 miliardi di anni fa, irrompendo nell'esistenza con una grande esplosione che continua ancora oggi. Nella versione più semplice della teoria, l'universo si espande ovunque uniformemente a partire dal punto singolare. L'uniformità di questa espansione spiega particolarmente bene gran parte dei dati osservativi: la materia extragalattica si allontana dalla nostra galassia a una velocità che varia in modo regolare con la distanza, e un bagno freddo di radiazione riempie tutto il cielo nella regione delle microonde, a una temperatura che, su angoli di qualche grado, varia di meno di una parte su 30 000. Nonostante questi successi, altri dati ci costringono ad ammettere che l'espansione non sia esattamente uniforme. Se così fosse, infatti, la materia non riuscirebbe ad aggregarsi e l'universo diventerebbe un gas di particelle elementari sempre più rarefatto: le stelle e le galassie non esisterebbero.

Per poter spiegare l'esistenza di una struttura nello stato attuale dell'universo, quindi, il cosmologo del big bang deve ammettere che l'universo presentasse anche prima una certa «granulosità». Queste prime disomogeneità potevano anche essere lievi e quasi indistinguibili dal fondo omogeneo; piccole fluttuazioni nella curvatura dell'universo primordiale avrebbero preso la forma di leggere compressioni e rarefazioni della materia e dell'energia da una regione all'altra dello spazio. L'ampiezza delle fluttuazioni (cioè la variazione rispetto alla densità media) doveva essere abbastanza grande da svilupparsi negli aggregati di materia oggi osservabili durante il tempo trascorso dall'inizio dell'universo; quanto grande esattamente debba però essere questa ampiezza è un problema molto delicato dal punto di vista teorico. Se le fluttuazioni iniziali fossero state troppo grandi avrebbero causato variazioni nella temperatura della radiazione di fondo a microonde che invece non sono visibili. Per di più, le fluttuazioni devono dare origine proprio alle strutture di scala abbastanza particolare che costituiscono l'universo, e non a strutture di grandezza arbitraria. Oggi è possibile identificare stelle, galassie, ammassi di galassie e perfino superammassi, cioè ammassi di ammassi, ma su masse di ordine superiore a quello dei superammassi l'universo appare piuttosto uniforme.

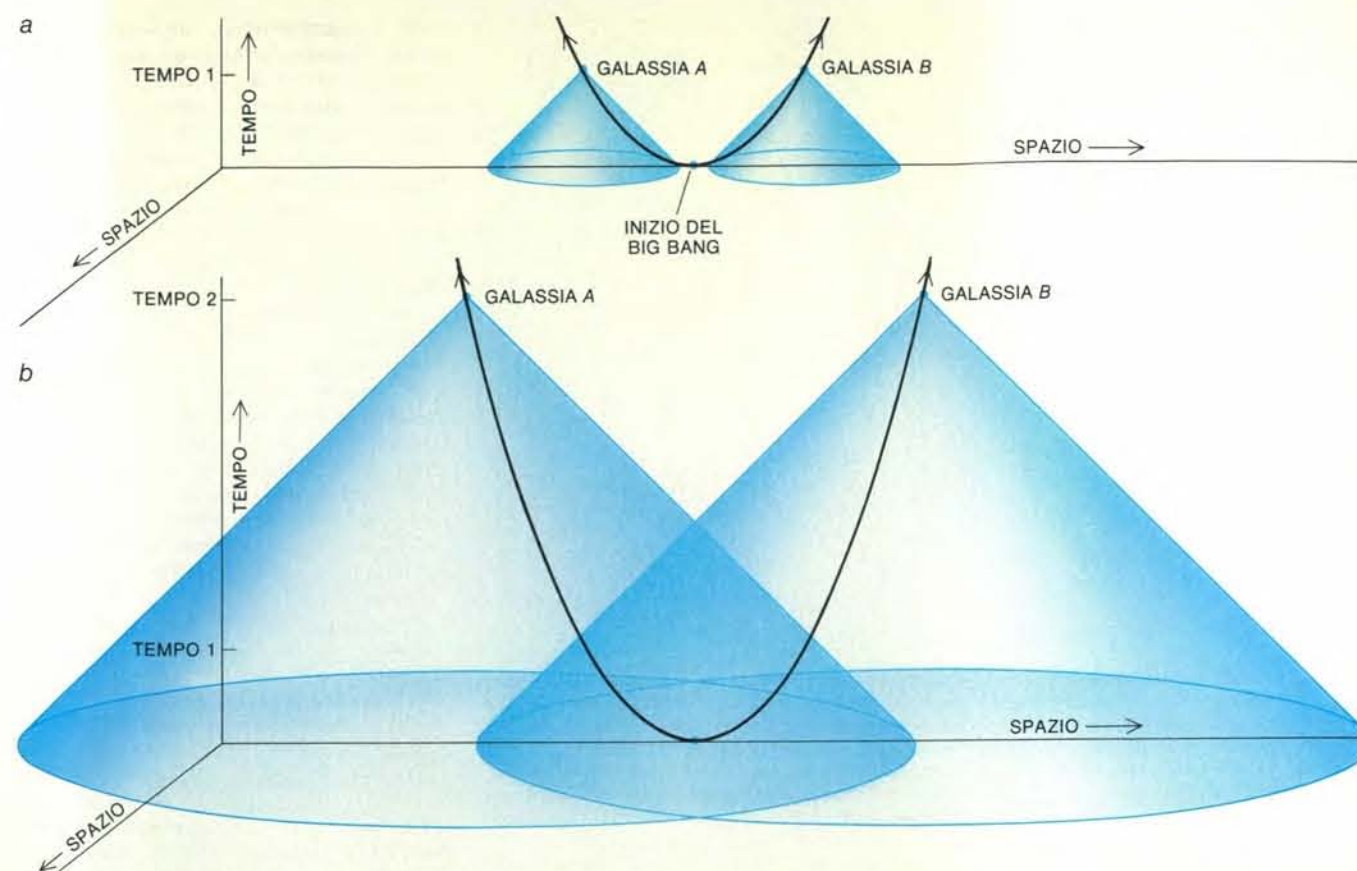
Negli ultimi tempi la convergenza tra cosmologia e fisica delle particelle ha permesso di soddisfare tutti questi requisiti senza formulare nessuna ipotesi restrittiva sullo stato dell'universo primordiale. In particolare non è necessario fare appello a scale o strutture particolari di massa e di energia all'avvio dell'espansione, né invocare nuove forze. Tutto ciò che si ipotizza è che subito dopo l'inizio del big bang esistessero ovunque nell'universo piccole variazioni nella densità della materia e dell'energia. Queste variazioni erano dovute alla sovrapposizione di fluttuazioni sinusoidali di piccola ampiezza nella densità a ogni possibile lunghezza d'onda, o ordine di grandezza delle lunghezze d'onda; le ampiezze delle fluttuazioni avevano una distribuzione casuale, così che le variazioni di densità risultanti erano casuali e caotiche. A partire da questa situazione la struttura attuale dell'universo si sarebbe potuta evolvere seguendo principi fisici abbastanza chiari.

Durante l'espansione dell'universo il moto libero casuale in tutte le direzioni delle particelle elementari cancellò tutte le fluttuazioni iniziali al di sotto di una dimensione critica: le uniche fluttuazioni sopravvissute furono quelle in cui venivano compresse o rarefatte masse pari ad almeno 10^{15} o 10^{16} volte quella del Sole. In seguito la gravità fece sì che alcune fra le masse compresse si contraessero soprattutto su uno dei tre assi spaziali. In questo modo lo spettro iniziale di fluttuazioni diede origine a gigantesche nubi irregolari di gas simili a frittelle appiattite. Dove queste si intersecavano si formarono filamenti di materia lunghi e sottili. Alcune delle nubi rimasero intatte, altre si frammentarono formando galassie e ammassi di galassie. L'emergere di una scala caratteristica opportuna delle fluttuazioni è stato spiegato per la prima volta da uno di noi (Silk), e un altro di noi (Zel'dovich) ha scoperto la formazione di strati sottili di materia per effetto della gravitazione. Nel seguito daremo a questo modello il nome di «teoria delle frittelle» (*pancake theory*).

La teoria delle frittelle nella sua forma attuale è la storia di due oggetti che si trovano ai due estremi della scala dei fenomeni fisici. Uno è un sistema astronomico così grande che riempie 10^{23} anni luce cubi di spazio; l'altro è il neutrino, una particella elementare che interagisce molto debolmente con la materia e che è quasi infinitamente piccola. Perché la teoria delle frittelle trovi una conferma, è necessario che vengano osservati entrambi questi oggetti e che al secondo si assegni una massa diversa da zero; tuttavia, dal momento che le loro masse, se esistono tutte e due, abbracciano un intervallo di 80 ordini di grandezza, per misurarle dalla nostra posizione privilegiata su una scala intermedia occorrono procedimenti straordinari.

Questa carta della distribuzione di 400 000 galassie su un'area del cielo di 100 gradi di diametro è stata realizzata da un programma per calcolatore studiato per intensificare la struttura filamentosa che si può percepire in altre carte. Il numero delle galassie contenute in ciascun pixel (elemento minimo dell'immagine) è indicato da un colore convenzionale: i pixel neri rappresentano aree con il minor numero di galassie, e le sfumature di marrone via via più chiare fino al bianco indicano densità di galassie crescenti. I pixel verdi e quelli rossi corrispondono a creste e picchi locali nella distribuzione: i pixel verdi indicano regioni in cui il numero delle galassie presenti è più grande che in tutti i pixel adiacenti lungo due o tre direzioni, orizzontale, verticale o diagonale. I pixel rossi invece indicano aree in cui il numero delle galassie raggiunge un massimo locale lungo tutte e quattro le direzioni. Le creste filamentose verdi forse corrispondono a superammassi di galassie addensatisi nello spazio; se è così, queste formazioni si estendono per 100 milioni di anni luce. La carta è stata realizzata sulla base di un rilevamento di C. Donald Shane e Carl A. Wirtanen del Lick Observatory: il centro della mappa è il polo nord galattico, nella costellazione Coma Berenices, il limite esterno è costituito dal parallelo galattico a 40 gradi nord, e l'orientamento è scelto in modo che il centro della galassia si trovi in alto. La carta è stata realizzata alla Princeton University da John E. Moody, Edwin L. Turner e J. Richard Gott III.

Poco dopo l'inizio del big bang, in tutto l'universo esistevano piccole perturbazioni nella densità della materia e dell'energia. Le perturbazioni possono essere viste come fluttuazioni, simili a onde, della densità attorno a un valore medio distribuite a caso su tutte le lunghezze d'onda. La parte superiore di questa immagine prodotta al calcolatore rappresenta un'istantanea di una di queste fluttuazioni, vista in sezione. Il giallo, il verde e il blu indicano regioni relativamente compresse, mentre l'arancione, il rosso e il viola corrispondono ad aree via via più rarefatte. Nella sezione inferiore dell'immagine tutte le fluttuazioni che coinvolgono masse inferiori a 10^{16} masse solari sono state filtrate ed eliminate dall'interazione tra materia e radiazione nelle prime fasi dell'evoluzione dell'universo. Secondo la *pancake theory*, ossia «teoria delle frittelle», proposta dagli autori, le fluttuazioni residue hanno dato origine ai superammassi di galassie e ai vuoti che questi hanno lasciati liberi, le formazioni che osserviamo attualmente. La simulazione al calcolatore è stata eseguita da S. Djorgovsky dell'Università della California a Berkeley.

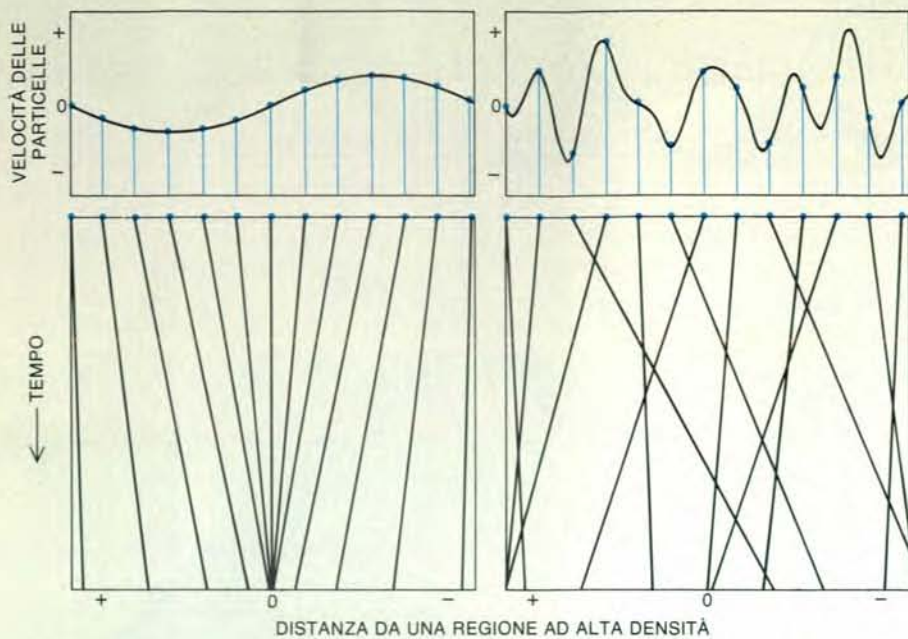
Negli ultimi tempi è stato verificato che esiste la scala del sistema astronomico richiesta e la sua massa sembra suggerire, anche se non basta a garantirlo, che la teoria delle frittelle è sulla buona strada. Sono state compiute misurazioni sistematiche delle distanze di diverse migliaia di galassie determinando lo spostamento verso il rosso dei loro spettri, lo spostamento cioè delle righe spettrali verso l'estremità dello spettro elettromagnetico corrispondente alle lunghezze d'onda maggiori. Lo spostamento verso il rosso è un effetto Doppler causato dall'allontanamento di una galassia lontana dalla nostra. La velocità di recessione si può calcolare dallo spostamento verso il rosso con una semplice formula matematica, e la distanza della galassia è direttamente proporzionale alla velocità di allontanamento perché l'universo è in espansione. Combinando una misurazione dello spostamento verso il rosso con le coordinate della galassia sulla sfera celeste, un astronomo può stabilire quindi la posizione della galassia nello spazio; in questo modo è stato possibile elaborare carte tridimensionali della distribuzione delle galassie.

Queste carte rivelano caratteristiche molto diverse da quelle di quasi tutti gli altri oggetti astronomici: le galassie sono concentrate in enormi strutture laminari e filamentose la cui dimensione massima, circa 100 milioni di anni luce, supera di un ordine di grandezza quella minima. Una struttura di questo genere può contenere fino a un milione di galassie, e la sua massa è dell'ordine delle 10^{16} masse solari. Inoltre le galassie non sono distribuite uniformemente all'interno della struttura: si distinguono grumi e nastri più densamente popolati, molti dei quali si trovano all'intersezione di due lamine. Infine, dispersi qua e là tra le strutture più grandi, si aprono enormi vuoti, praticamente privi di galassie, con dimensioni tra i 100 e i 400 milioni di anni luce (si veda l'articolo *Superammassi e vuoti nella distribuzione delle galassie* di Stephen A. Gregory e Laird A. Thompson in «Le Scienze», n. 165, maggio 1982). Quest'immagine dell'universo si basa in buona parte sul lavoro di Jaan Einasto dell'Osservatorio di Tartu in Estonia.

L'individuazione di un neutrino dotato di massa è molto più problematica. Parecchi anni fa i fisici teorici attribuirono al neutrino una massa di riposo nulla, ma alcune teorie più recenti sulle particelle elementari fanno pensare che il neutrino in realtà abbia una piccola massa. Per rivelarla sono in corso diversi tipi di esperimenti: secondo il metodo più diretto è possibile dedurre l'esistenza di questa massa trovando certe variazioni particolari nel tasso di decadimento degli isotopi radioattivi. Nel 1980 Valentin Lubimov, Evgeny Tretyakov e loro collaboratori dell'Istituto di fisica teorica e sperimentale di Mosca hanno misurato il tasso di decadimento del trizio, l'isotopo radioattivo dell'idrogeno. All'epoca comunicarono risultati in accordo con una massa piccola, ma non nulla, per il neutrino, tra i 14 e i 46 elettronvolt, cioè meno di un decimillesimo della massa dell'elettrone. Negli ultimi tempi gli stessi ricercatori hanno confermato quanto avevano ottenuto precedentemente, restringendo inoltre i margini di errore: oggi infatti attribuiscono alla massa del neutrino un valore che è compreso tra i 20 e i 40 elettronvolt.

L'orizzonte di un osservatore si allarga con il passare del tempo, abbracciando una parte sempre più grande dell'universo. L'orizzonte è una sfera centrata sull'osservatore e di raggio pari alla distanza che la luce ha potuto percorrere dall'inizio del big bang al momento considerato. Qui è rappresentato con un cerchio alla base di un cono: l'espansione dell'universo a partire dall'inizio del big bang è indicata dal progressivo allontanamento di due galassie con il tempo. Nelle prime fasi dell'espansione le galassie si allontanano l'una dall'altra con una velocità apparente superiore a quella della luce; quindi in quest'epoca per un osservatore in una di esse non c'è stato abbastanza tempo per vedere l'altra in nessun momento precedente della sua evoluzione (*a*). Ma, poiché l'espansione rallenta continuamente, la luce emessa da una galassia all'inizio del suo sviluppo finirà per raggiungere un osservatore situato sull'altra (*b*). La superficie dei coni è costituita dalle traiettorie dei segnali luminosi nello spazio e rappresenta un limite spaziale all'interazione causale in qualsiasi momento dato, perché nessun segnale può propagarsi a velocità superiore a quella della luce. Per questo, qualunque fluttuazione di densità che si verifichi su scala più grande di quella dell'orizzonte di un osservatore non avrà alcun effetto sull'osservatore o sulla distribuzione della materia e dell'energia entro l'orizzonte.

Sfortunatamente non si sono avute verifiche indipendenti di questi risultati; e pertanto manca ancora un accordo generale sulla questione della massa del neutrino. Un secondo tipo di esperimento, tentato per la prima volta da Ettore Fiorini dell'Università di Milano, si basa sulla velocità di un certo processo di decadimento radioattivo, noto come doppio decadimento beta, visibile in certi isotopi. Fiorini ha riferito che la massa del neutrino non può essere maggiore di 10 o 20 elettronvolt, dallo studio del decadimento dell'isotopo germanio 76. Il metodo è meno diretto della misurazione del decadimento del trizio; i risultati dell'esperimento di Fiorini, infatti, si possono interpretare come una misura della massa del neutrino solo nell'ipotesi che il neutrino sia l'antiparticella di se stesso. Se invece il neutrino e l'antineutrino sono particelle distinte, il doppio decadimento beta del germanio 76 è diverso, e non è possibile dedurne alcun valore per la massa del neutrino.

Un terzo metodo per rivelare la massa dei neutrini è stato proposto per la prima volta da Bruno M. Pontecorvo dell'Istituto unito di ricerche nucleari di Dubna, in Unione Sovietica. Il metodo sfrutta il fatto che esistono tre diverse specie di neutrino: il neutrino elettronico, il neutrino muonico e il neutrino tauonico. Se tutte e tre le specie sono dotate di massa, se possono manifestarsi con probabilità variabile e se la differenza tra i quadrati delle masse di due neutrini di specie diverse è diversa da zero, allora, per la meccanica quantistica, le tre specie possono oscillare, cioè scambiarsi liberamente le loro identità. Dal momento che queste oscillazioni farebbero variare nel tempo la popolazione di ciascuna specie, dovrebbe essere possibile rivelarle come variazioni nella popolazione di neutrini elettronici, per esempio, in un fascio di neutrini dei tre tipi. Negli ultimi anni sono stati eseguiti diversi esperimenti di questo genere, prima dal gruppo di Frederick Reines dell'Università della California a Irvine nel 1980, e poi dal gruppo di Felix H. Boehm al California Institute of Technology e da altri ricercatori. Fino alla stesura di questo articolo nessun gruppo sperimentale ha comunicato prove univoche dell'esistenza di oscillazioni. Sfortunatamente la loro assenza potrebbe indicare anche solo che la differenza tra i quadrati delle masse di due specie di neutrini è nulla; anche un insuccesso nella rivelazione delle oscillazioni è compatibile con una massa finita, cioè non nulla, del neutrino.

L'atteggiamento prevalente tra i fisici è che i risultati sperimentali non permettano ancora una conclusione sicura sul problema della massa del neutrino. Se però si accetta la validità delle prove a favore della sua esistenza, le conseguenze cosmologiche sono molto profonde. Più di dieci anni fa, seguendo un suggerimento precedente di Semyon Gershtein dell'Istituto di fisica di Serpukhov, in Unione Sovietica, e di uno di noi (Zel'dovich), György Marx e un altro di noi (Szalay)

Una fluttuazione con un'unica lunghezza d'onda genera una distribuzione regolare delle velocità delle particelle; le particelle vengono attirate dalla gravità verso le regioni di densità massima, quindi la velocità di ognuna di esse dipende dalla sua distanza da una regione di questo tipo (*in alto a sinistra*). La velocità di ogni particella è data dalla pendenza della linea che ne rappresenta la traiettoria (*in basso a sinistra*). Le traiettorie tendono a convergere e a formare regioni di maggiore densità, amplificando così le fluttuazioni regolari. Se invece le fluttuazioni si distribuiscono casualmente su tutte le lunghezze d'onda (*in alto a destra*) le traiettorie non convergono (*in basso a destra*).

all'Università Eötvös ipotizzarono che i neutrini dotati di massa potessero dare un contributo molto importante alla massa e all'evoluzione complessiva dell'universo. Lo stesso suggerimento fu avanzato contemporaneamente da Ramanath Cowsik e John McClelland dell'Università della California a Berkeley. Un neutrino dotato di massa porterebbe inoltre necessariamente alla formazione di strutture a frittella sulle scale più grandi. Prima di discutere questo effetto, però, sarà utile descrivere una versione precedente della teoria delle frittelle, una teoria che alla fine ha fallito diverse verifiche osservative cruciali, ma che ha dato vita alla formulazione attuale, di maggior successo.

Gli astrofisici ritengono di aver compreso piuttosto bene i processi fisici che debbono essersi verificati dopo i primi millisecondi dal big bang. Le energie delle particelle che collidevano tra di loro in quell'epoca non superavano quelle che si raggiungono normalmente nei piccoli acceleratori; per costruire quindi un quadro dell'universo di allora basta studiare un fluido denso costituito da particelle le cui proprietà individuali siano ben note.

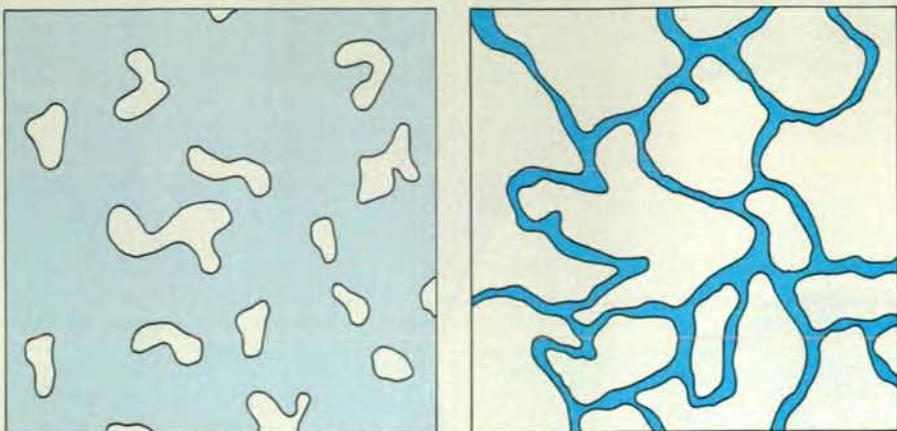
Le particelle di gran lunga più abbondanti in quel fluido erano il fotone, l'elettrone e le tre specie di neutrino; i protoni e i neutroni sopravvissuti alle annichilazioni del periodo precedente erano relativamente pochi. Gli elettroni e i neutrini rimasero in stretto contatto per quasi tutto il primo secondo, annichilandosi e riformandosi continuamente. La frequenza delle collisioni garantiva che l'energia fosse distribuita in modo casuale in tutto il fluido; le particelle, in altre parole, erano mantenute in equilibrio termico. Durante la progressiva espansione dell'universo la densità delle particelle diminuì, e le collisioni si fecero meno frequenti. Dato che l'energia di una particella è inversamente proporzionale alla sua lunghezza d'onda, l'energia media dei fotoni diminuì mentre le loro lunghezze d'onda si espandevano con il resto dell'universo, che cominciò così a raffreddarsi.

Oggi le ricerche teoriche che negli ultimi tempi hanno tentato di unificare le forze della natura sono in grado di risalire nella storia dell'universo anche oltre il primo millisecondo. Queste teorie si chiamano di grande unificazione perché si sforzano di interpretare la forza elettromagnetica, quella nucleare debole e quella nucleare forte come manifestazioni diverse a bassa energia di un unico fenomeno sottostante. (La gravità, la quarta forza fondamentale, non è stata ancora incorporata in questo schema.) La densità di energia alla quale le tre forze non sono più distinte corrisponde a quella dell'universo solo 10^{-35} secondi dopo l'inizio del big bang. I primissimi istanti dell'universo sono diventati quindi agli occhi degli scienziati un laboratorio per mettere alla prova le previsioni delle teorie di grande unificazione.

Una delle previsioni di queste teorie è che se la densità della materia nelle prime fasi dell'espansione deve fluttuare, deve fluttuare anche quella dei fotoni, cioè della radiazione. Nonostante questo, però, il rapporto tra le due densità deve rimanere sempre uguale. Secondo la teoria della relatività generale la materia e l'energia sono equivalenti come sorgenti gravitazionali e determinano la geometria dello spazio-tempo. Una fluttuazione nella densità della materia e dell'energia, quindi, causa una fluttuazione del campo gravitazionale, equivalente a sua volta a una fluttuazione nella curvatura dello spazio-tempo. La teoria generale di queste fluttuazioni nell'universo in espansione, trattate nel quadro della relatività generale, è stata elaborata nel 1946 da Eugene M. Lifshitz dell'Istituto per i problemi fisici di Mosca.

Sembra ragionevole supporre che nell'universo, all'inizio, sia esistita una vasta gamma di fluttuazioni su diverse scale possibili. Facciamo questa ipotesi soprattutto per economia di pensiero: sembrerebbe arbitrario e completamente fortuito che le fluttuazioni fossero state tali da individuare solo regioni su una certa scala, poniamo su scala galattica. Esiste però un limite superiore alle dimensioni delle fluttuazioni che in un dato momento un osservatore può percepire. Questo limite è l'orizzonte spaziale dell'osservatore, una sfera centrata sull'osservatore stesso di raggio pari alla distanza che la luce può percorrere nel tempo trascorso dall'inizio del big bang. Nel modello standard del big bang, comunque, l'espansione iniziale dello spazio e del tempo a partire dal punto singolare crea un universo molto più grande dell'orizzonte spaziale di qualsiasi osservatore. D'altra parte, dato che si ritiene che l'espansione dell'universo stia rallentando, nell'orizzonte di ogni osservatore entrano quantità di materia sempre più grandi. Le fluttuazioni che non sono rilevabili a un certo stadio dell'evoluzione dell'universo lo diventano più tardi, perché cominciano a essere contenute all'interno dell'orizzonte in continua estensione (*si veda l'illustrazione nella pagina precedente*).

Una volta all'interno dell'orizzonte di un'osservatore, una fluttuazione può essere descritta in maniera adeguata entro la teoria della gravitazione classica, ossia non relativistica. In questa teoria l'oscillazione assume l'aspetto di una perturbazione osservabile della densità del fluido. Su ogni piccola porzione di materia e di radiazione che si muovono insieme agiscono due effetti contrastanti: quello della gravità che tende a farlo collassare e quello della pressione dovuta al moto caotico delle particelle e della radiazione che tende a disperderla nello spazio. Sulle grandi scale la gravità finisce sempre per prevalere: la pressione non riesce a impedire il collasso, e le particelle vengono attirate verso le regioni di densità massima. Per di più, una volta che il collasso gravitazionale è iniziato, la massa che si è andata accumulando attira materia e radiazione ancora più lontane, amplificando qualunque instabilità iniziale; la materia si accumula in certe regioni e diventa più rarefatta in altre.

Se si considera il brodo di particelle e radiazione di cui all'inizio è costituito l'universo come un gas ideale, gli effetti di una fluttuazione di densità sul gas sono

La struttura topologica delle fluttuazioni che sopravvivono alla selezione nelle prime fasi dell'evoluzione dell'universo si conserva nella struttura dei superammassi di galassie e dei vuoti intergalattici. A sinistra la densità dell'universo è quasi uniforme. Se la probabilità che una leggera fluttuazione nella densità causi una rarefazione lungo un asse qualsiasi è di un mezzo, la probabilità che in una certa regione avvenga una rarefazione su tutti e tre gli assi spaziali è di un ottavo. Le regioni in cui questo avviene sono rappresentate in bianco; inizialmente occupano solo un ottavo del volume dell'universo e contengono circa un ottavo della sua massa (*a sinistra*). La gravità comprime le regioni in colore lungo uno o più assi. Queste regioni, che racchiudono circa sette ottavi della massa dell'universo, collassano formando una ragnatela di filamenti che al termine del processo finiscono per occupare solo un ottavo del volume (*a destra*). Le regioni di rarefazione iniziale si espandono formando vuoti che occupano i rimanenti sette ottavi dell'universo.

facili da calcolare. Qualsiasi aumento di densità locale che coinvolga una massa abbastanza grande porterà all'instabilità gravitazionale e all'avvio del collasso. Su scale più piccole, invece, la gravità non è abbastanza forte da superare l'aumento di pressione causato dalla maggiore densità del gas. La porzione di gas compressa rimbalzerà, quindi, rarefacendosi, e la fluttuazione si propagherà esattamente come un'onda sonora, cioè sotto forma di compressioni e rarefazioni periodiche del mezzo nel quale si sposta.

Quasi tutte le onde sonore nell'aria si estinguono in qualche decina di metri perché le particelle che costituiscono le onde di compressione si disperdono e il loro moto coerente viene dissipato sotto forma di calore. Analogamente le onde sonore nel mezzo cosmico generate dalle fluttuazioni perdono la loro energia e si smorzano a tutte le lunghezze d'onda tranne le più grandi. Inoltre le particelle e i fotoni dell'universo primitivo sono di gran lunga troppo addensati per poterli trattare come un gas ideale. Nei primi 300 000 anni del big bang la radiazione fotonica possedeva abbastanza energia da mantenere ionizzata tutta la materia. I fotoni erano circa 100 milioni di volte più numerosi degli elettroni; gli elettroni liberi, che più tardi si sarebbero legati ai nuclei, erano sottoposti quindi a un bombardamento continuo da parte dei fotoni, con fenomeni di diffusione reciproca. Il risultato era un fluido di elettroni e di fotoni denso e viscoso, in cui le particelle potevano percorrere solo piccole distanze prima di subire una diffusione.

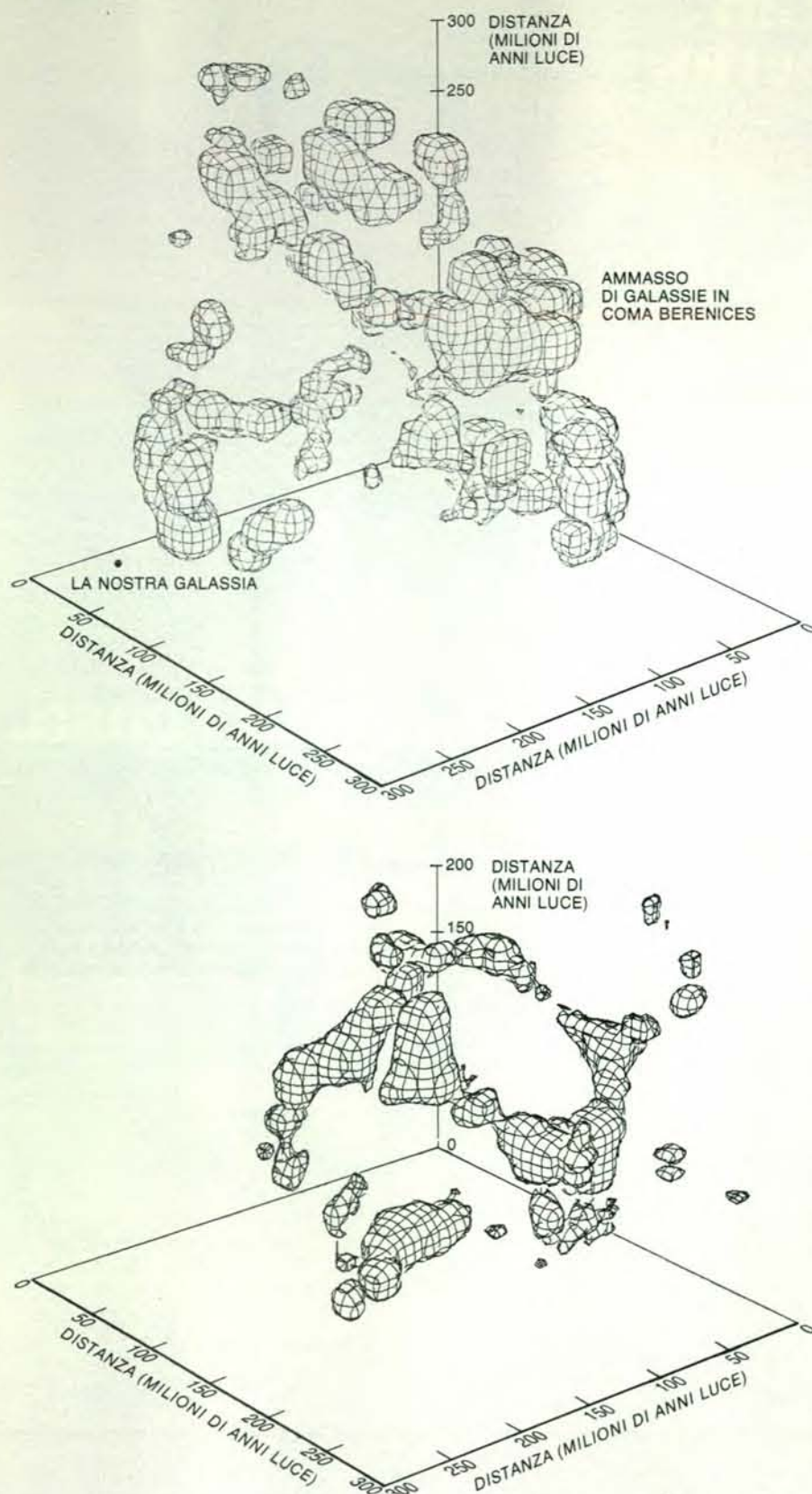
La diffusione degli elettroni liberi da parte della radiazione rende il loro spostamento attraverso di essa molto simile all'attraversamento di una specie di miele viscoso e freddo. La viscosità del fluido impedisce quindi lo sviluppo di quelle instabilità gravitazionali che potrebbero derivare dalla sola aggregazione di materia. Inoltre, come nel gas ideale, l'intensità della pressione di radiazione è tale da impedire alla materia e all'energia di collassare sotto l'azione della gravità, tranne che su regioni di scala abbastanza grande. Tutte le altre fluttuazioni nel fluido viscoso, cioè quelle che sfuggono al collasso gravitazionale, si possono considerare onde sonore.

Come abbiamo già ricordato, le teorie di grande unificazione richiedono che il rapporto tra densità dei fotoni e densità della materia rimanga costante: nella fase compressiva di una fluttuazione, quindi, deve avvenire una compressione dei fotoni pari a quella delle particelle dotate di massa. Se però la distanza percorsa da un fotone dall'inizio del big bang al momento considerato è maggiore della lunghezza di una regione compressa di una fluttuazione, il fotone in effetti non prenderà parte alla compressione, ma dissiperà la frazione che gli compete dell'energia della fluttuazione. I fotoni, molto più numerosi delle particelle dotate di massa, trasportano pressoché tutta l'energia delle fluttuazioni, quindi tutte le fluttuazioni su scala inferiore allo spostamento radiale di un fotone durante il tempo trascorso dall'inizio del big bang vengono smorzate completamente.

Il cammino di un fotone è simile a quello di un ubriaco che, allontanandosi barcollando da un lampione, prende a ogni passo qualsiasi direzione con la stessa probabilità. Per allontanarsi dal lampione di una distanza pari a N passi di un uomo sobrio l'ubriaco deve fare N^2 passi. Analogamente un fotone deve subire N^2 diffusioni per percorrere una distanza radiale pari a quella che percorrerebbe muovendosi liberamente (nel tempo necessario a subire N diffusioni). Nonostante siano diffusi dagli elettroni, i fotoni si disperdono attraverso il mezzo tanto rapidamente da dissipare l'energia di tutte le fluttuazioni, tranne le più grandi. Quando l'universo si è raffreddato abbastanza da permettere ai nuclei atomici di catturare gli elettroni, i fotoni si sono espansi attraverso una regione di massa pari a circa 10^{14} volte quella del Sole, e tutte le fluttuazioni iniziali che coinvolgono masse inferiori a questo limite sono state cancellate.

Quando alla fine gli elettroni si combinano con i nuclei formando atomi, la materia e la radiazione cessano di interagire, e la radiazione comincia a muoversi liberamente, indipendentemente dagli atomi. La viscosità del fluido scende improvvisamente a un valore molto più basso e le fluttuazioni che sono sopravvissute all'era precedente di interazioni dominate dalla radiazione possono venire amplificate senza più ostacoli. Da questo momento l'instabilità gravitazionale procede a pieno ritmo.

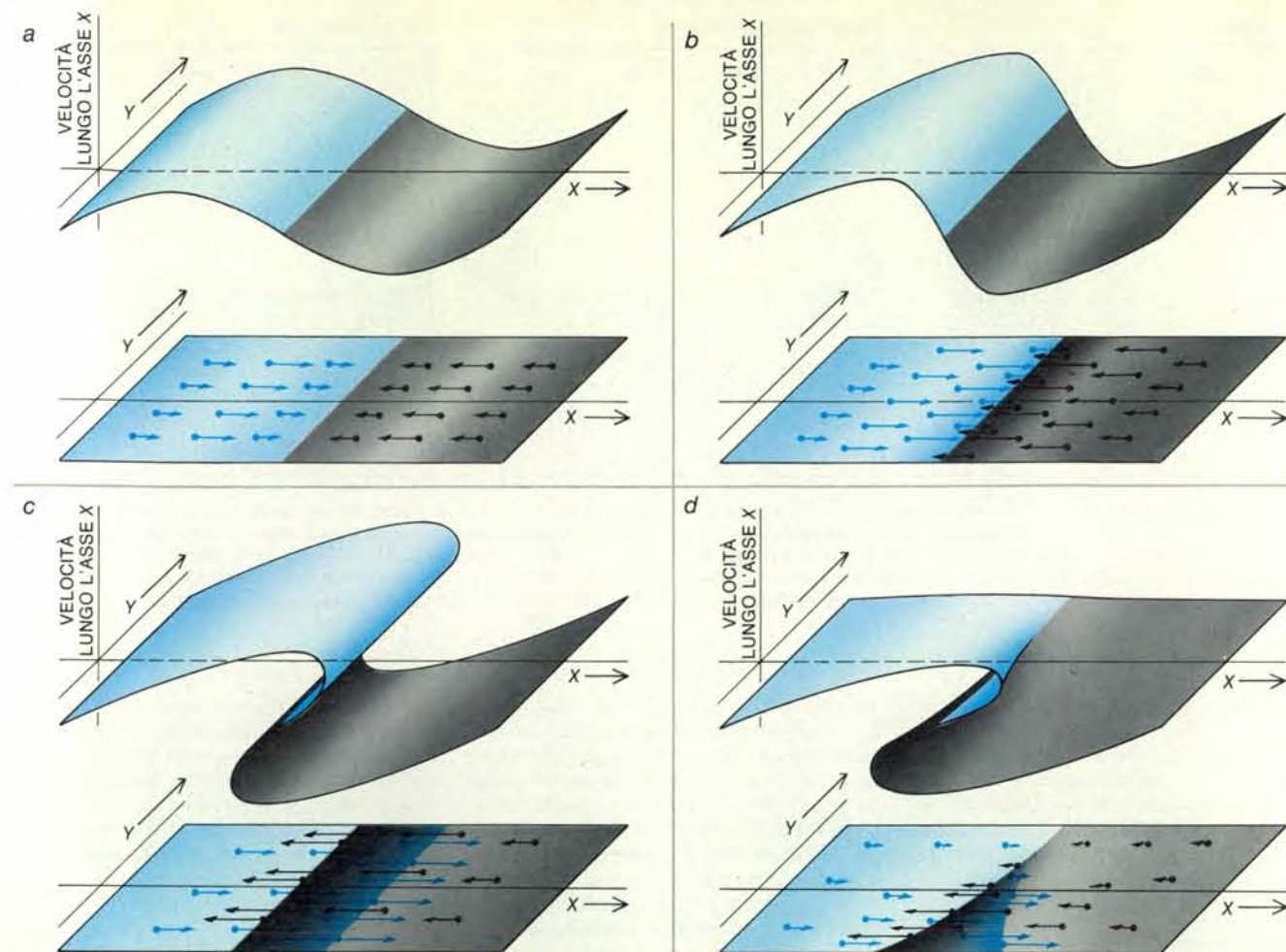
L'improvvisa scomparsa della pressione di radiazione ha un effetto determinante per stabilire la forma e la struttura degli oggetti che si sviluppano per primi. La pressione termica agisce sempre in maniera isotropa, cioè nello stesso modo in tutte le direzioni, quindi se la pressione di radiazione fosse rimasta di intensità paragonabile alla forza di gravità tutti gli oggetti collassati avrebbero avuto una simmetria sferica quasi perfetta. Le anisotropie si sviluppano perché la pressione rimane trascurabile fino agli ultimi istanti del collasso.

Data la mancanza di una pressione che si opponga alla contrazione, la instabilità gravitazionale si rivela un meccanismo molto efficiente per radunare la quasi totalità della materia in regioni dello spazio compresse, di alta densità. Si consideri questo ragionamento: lungo ciascuno dei tre assi spaziali la materia può essere compressa oppure rarefatta; supponiamo, per semplicità, che la probabilità che la materia sia compressa lungo un certo asse sia di un mezzo. Allora la frazione del gas totale non compressa su nessuno dei tre assi è il cubo di un mezzo, cioè un ottavo. Il risultato ha conseguenze immediate sulla previsione della struttura spaziale successiva al collasso. In una prima fase, quando la densità è ancora quasi uniforme, le regioni che verranno poi compresse contengono circa i sette ottavi della materia. Queste regioni circondano bolle più piccole di materia che non collasserà mai: le bolle formeranno i vuoti. Dopo il collasso le regioni compresse occupano solo un ottavo del volume totale; le piccole bolle, che contengono solo un ottavo della massa, si espandono riempiendo i rimanenti sette ottavi del volume. La topologia dello stato iniziale si conserva. Il risultato finale è una struttura cellulare formata da pareti sottili e filamenti di materia compressa che racchiudono grandi vuoti.

Le isolinee di densità elevata sono state riportate in un grafico tridimensionale relativamente a tutte le galassie più luminose della magnitudine 14,5 che nel cielo settentrionale si trovano entro circa 250 milioni di anni luce dalla nostra galassia (*illustrazione in alto*). La distribuzione è simile a quella che compare nell'illustrazione in basso e che è stata ottenuta simulando al calcolatore le previsioni del modello a frittelle. Nel modello si suppone che il raggrupparsi della materia in ammassi derivi solo da fluttuazioni della densità non più piccole della più piccola fluttuazione in grado di sopravvivere agli effetti di smorzamento di eventuali neutrini dotati di massa. La distribuzione reale è stata ricavata da Carlos Frenk e Simon White dell'Università della California a Berkeley sulla base di una mappa preparata da Marc Davis e dai suoi colleghi del Center for Astrophysics dello Harvard College Observatory e dello Smithsonian Astrophysical Observatory. La mappa si basava su un rilevamento di 1801 galassie. La simulazione è stata eseguita da Frenk, White e Davis.

Con considerazioni analoghe è possibile prevedere anche la forma delle regioni compresse: è estremamente improbabile che un volume cubico di materia destinato al collasso formi una sfera. Per questo sarebbe necessario che fossero uguali, su tutte e tre le componenti in cui si può risolvere qualunque collasso, sia la direzione, sia l'entità delle fluttuazioni. È molto più probabile che il cubo collassi innanzitutto lungo un asse scelto a caso, e collassi o si espanda più lentamente lungo gli altri due. La distribuzione della materia che ne segue è estremamente anisotropa: la massa contenuta nel volume cubico iniziale non varia al diminuire dello spessore e del volume, quindi la densità diventa molto elevata e si forma una specie di frittella piatta.

All'inizio le frittelle si sviluppano in regioni isolate, ma presto crescono diventando lamine sottili che si intersecano e formano la struttura cellulare. Alcune simulazioni numeriche del collasso eseguite con l'aiuto di grandi calcolatori fanno pensare che l'universo abbia assunto una struttura cellulare solo in tempi recenti. Nel futuro, via via che si formeranno aggregati più grandi di materia, si prevede che la struttura cellulare scomparirà. Quindi è solo durante uno stadio intermedio dell'evoluzione del cosmo che la struttura della distribuzione di materia rispecchia le fluttuazioni iniziali della curvatura. I dati osservativi mostrano che, dal punto di vista dell'evoluzione della struttura su grande scala, l'universo non è né molto giovane né molto vecchio.

Esistono due grandi difficoltà nella teoria delle frittelle come l'abbiamo descritta finora. Innanzitutto, come si ricorderà, in questa teoria le più piccole fluttuazioni in grado di sopravvivere all'era della radiazione coinvolgono 10^{14} masse solari. Nella distribuzione delle galassie, però, esiste una struttura su scale molto più grandi. Le simulazioni numeriche sembrano favorire una teoria in cui le più piccole fluttuazioni che appaiono alla fine dell'era della radiazione siano dell'ordine di 10^{15} o 10^{16} masse solari.

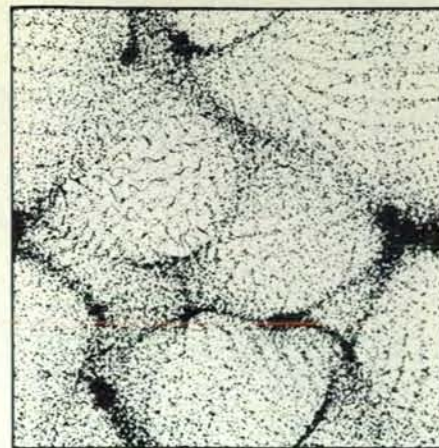
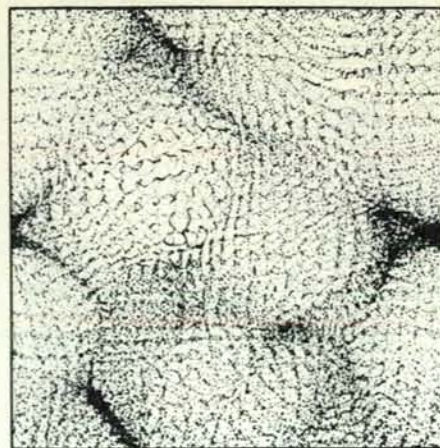
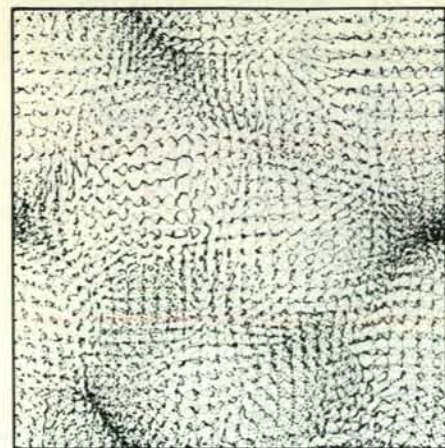
La seconda difficoltà è più grave. Il fondo di radiazione a microonde ha continuato a propagarsi liberamente fin dal momento in cui i fotoni e gli elettroni hanno smesso di interagire; per questo la variazione di temperatura della radiazione tra un punto e l'altro del cielo rispecchia le disomogeneità primordiali nella distribuzione della materia. All'epoca della formulazione della teoria originale delle frittelle il limite superiore per la variazione di temperatura complessiva su tutto il cielo era di circa una parte su mille, e di conseguenza si riteneva che le disomogeneità primordiali della materia nell'universo potessero arrivare a un terzo della variazione di temperatura, cioè a una parte su 3000. Di recente Francesco Melchiorri e i suoi collaboratori alle Università di Firenze e di Roma e Yuri N. Parijskijin dell'Osservatorio di Pulkovo a Leningrado hanno stabilito limiti molto più stretti per le variazioni della temperatura della radiazione: il nuovo limite superiore è una variazione di meno di una parte su 30 000 su un angolo di sei gradi.

Le deformazioni continue del mezzo durante il collasso gravitazionale possono portare a discontinuità nella densità della materia. In ogni quadro di questa sequenza schematica il piano inferiore rappresenta due direzioni nello spazio; per semplicità si tiene conto solo dei moti paralleli all'asse *x* di particelle contenute nel piano. Le particelle si muovono verso un asse al centro del piano; la lunghezza di ogni freccia indica la velocità di una particella che si trovi alla base della freccia (*a*, *b*). Una «lamina» di particelle in moto verso destra (*frecce in colore*) attraversa l'asse centrale senza scontrarsi con le particelle che si spostano verso sinistra (*frecce in nero*). I due moti stabiliscono due discontinuità nella densità del mezzo ai lati dell'asse centrale (*c*). In quasi tutti i collassi che avvengono nella realtà i moti delle due lamine di particelle non sono esattamente simmetrici e le due discontinuità finiscono in una cuspide (*d*). Queste distribuzioni di densità possono costituire un caso particolare di un fenomeno più generale descritto dalla branca della matematica che si chiama teoria delle catastrofi. Rappresentando i moti delle particelle in uno spazio delle fasi, cioè in uno spazio tridimensionale in cui l'asse verticale rappresenta la velocità in una direzione parallela all'asse *x*, l'interazione tra le lamine di particelle è raffigurata da una superficie attorcigliata o ondulata (*piano superiore in ogni riquadro*). La densità delle particelle in ciascun punto è data, allora, dall'«ombra» che la superficie deformata getta sul piano *x*-*y* originale.

Le fluttuazioni richieste dalla versione originale della teoria delle frittelle erano in accordo con le stime precedenti delle variazioni di temperatura, ma con le successive c'è solo una corrispondenza limitata. Inoltre, se la densità complessiva di tutta la materia e l'energia dell'universo è tanto piccola da permettere all'espansione di continuare per sempre, l'accordo tra teoria ed esperienza scompare del tutto. Su scala cosmica la forza di gravità nel periodo più vicino a noi sarebbe stata così debole, che le fluttuazioni avrebbero dovuto svilupparsi completamente e collassare molto prima, quando la densità della materia era maggiore. L'ampiezza di queste fluttuazioni, però, sarebbe stata di gran lunga troppo grande per essere compatibile con l'uniformità del fondo di microonde. Invece, se le fluttuazioni iniziali fossero state abbastanza piccole da essere compatibili con il fondo di radiazione, la nascita delle galassie sarebbe diventata praticamente impossibile.

Se l'universo fosse abbastanza denso da permettere un accordo, per quanto limitato, tra l'ampiezza delle fluttuazioni e l'uniformità della radiazione di fondo, nascerebbe un altro problema. Non sarebbe possibile spiegare questa densità solo con la massa totale della materia luminosa visibile sotto forma di stelle, nebulose, galassie e simili; al contrario, l'universo devrebbe essere costituito soprattutto da materia oscura. Questa deduzione non è nuova: studi condotti sulla rotazione della nostra e di altre galassie a spirale hanno mostrato che le velocità di rotazione delle stelle alla periferia di una galassia non sono in accordo con le leggi di Keplero. Queste leggi affermano che le velocità di rotazione dovrebbero diminuire al crescere della distanza dal centro della galassia, proprio come la velocità orbitale dei pianeti diminuisce al crescere della loro distanza dal Sole. Le stelle alla periferia delle galassie, invece, non sono più lente; le loro velocità di rotazione sono all'incirca costanti e indipendenti dalla distanza dal centro galattico. P. James E. Peebles e Jeremiah P. Ostriker della Princeton University ed Einasto hanno avanzato contemporaneamente l'idea che il dilemma si risolverebbe se la maggior parte della massa delle galassie a spirale fosse costituita da aloni di materia invisibile. Un ragionamento indiretto fa pensare che forse la materia oscura è presente in quantità ancora maggiori all'interno dei gruppi e degli ammassi di galassie. Questi sistemi si separerebbero nei loro componenti in un tempo estremamente breve se non fosse per l'attrazione gravitazionale della materia oscura. Si calcola che la materia oscura rappresenti forse il 90 per cento della massa dell'universo.

TRE MILIARDI DI ANNI

CINQUE MILIARDI DI ANNI

NOVE MILIARDI DI ANNI

Questa simulazione al calcolatore mostra l'evoluzione delle fluttuazioni su grande scala della densità di materia e di energia nel tempo, se si parte dal presupposto che esistano neutrini con massa non nulla. In conseguenza del collasso gravitazionale si sviluppano frittelle e filamenti, mentre la materia delle altre regioni dello spazio diventa sempre più rarefatta. La struttura che ne risulta assomiglia alla distribuzione di superammassi e vuoti visibile oggi. In parecchie regioni esistono cuspidi e altri tipi di discontinuità nella densità, riconoscibili e classificabili con il ricorso alla teoria delle catastrofi. La simulazione è stata realizzata da George Efstathiou dell'Università di Cambridge.

Per salvare dal naufragio la teoria delle frittelle occorreva urgentemente un nuovo costituente dell'universo, ed era necessaria una forma di materia oscura per spiegare i moti delle galassie. Un candidato naturale per entrambi i ruoli era il neutrino, anche se certe altre particelle esotiche, ma ancora non rivelate, come il fotino o il gravitino dotati di massa, avrebbero potuto assolvere lo stesso compito cosmologico. Le teorie delle particelle elementari prevedono che nel primo millisecondo del big bang esistesse un grande assortimento di particelle che interagivano debolmente in equilibrio termico. Molte di queste particelle potrebbero essere sopravvissute fino a oggi e, se fossero stabili, potrebbero avere profonde conseguenze per la cosmologia. Dato che la massa del neutrino si può misurare sperimentalmente, ci riferiremo d'ora in avanti al neutrino; ciononostante, anche se il neutrino risultasse privo di massa, la teoria delle frittelle non risulterebbe confutata.

Si ricorderà che nel primo secondo del big bang il brodo primordiale conteneva moltissimi neutrini. Anche oggi il rapporto tra fotoni e neutrini delle tre varietà insieme è solo di 11 a 9. I neutrini, a differenza dei protoni, degli elettroni e perfino dei fotoni, interagiscono con le altre particelle così debolmente che iniziano a muoversi liberamente molto prima dei fotoni. Per questo i neutrini, che all'inizio viaggiano alla velocità della luce, possono percorrere nelle prime fasi dell'evoluzione dell'universo distanze maggiori dei fotoni. Al termine dell'era della radiazione i neutrini avranno quindi dissipato fluttuazioni su scala più grande di quelle che sarebbero state smorzate dai soli fotoni.

Un neutrino dotato di massa non può continuare indefinitamente a muoversi alla velocità della luce. Quando la densità di energia dei fotoni scende al di sotto dell'energia che corrisponde approssimativamente alla massa di riposo del neutrino, questo comincia a rallentare e a muoversi a una velocità appropriata alla sua energia. Se la massa del neutrino è di 30 elettronvolt, il rallentamento comincerà molto prima della cattura degli elettroni da parte dei nuclei atomici. Per la cattura bisogna aspettare finché l'energia di fondo si riduce a 0,1 elettronvolt, l'energia alla quale l'idrogeno viene ionizzato dal denso fluido di fotoni. I neutrini continuano a smorzare le fluttuazioni anche mentre rallentano, ma diventano sempre più suscettibili alla cattura da parte di grandi fluttuazioni non ancora livellate. Richard Bond della Stanford University e uno di noi (Szalay) hanno valutato il massimo ordine di distanze che i neutrini possono percorrere liberamente prima di venire intrappolati e, quindi, il minimo ordine di grandezza al di sopra del quale le fluttuazioni non vengono cancellate. Questa scala corrisponde a una distanza attuale di 100 milioni di anni luce e a una massa di 10^{15} - 10^{16} volte quella del Sole. L'accordo con le dimensioni e la massa dei superammassi di galassie visibili oggi è impressionante.

Come si conciliano queste fluttuazioni con l'uniformità osservata della radiazione di fondo? I neutrini cessano di cancellare le fluttuazioni della curvatura prima della fine dell'era della radiazione, ma, a differenza degli elettroni, i loro moti non sono ostacolati dalla viscosità del fluido. I neutrini collidono con i fotoni e gli elettroni così raramente da non essere soggetti all'attrito viscoso. Per questo le instabilità gravitazionali possono iniziare a svilupparsi tra i neutrini prima della fine dell'era della radiazione, e quindi crescere per un tempo molto più lungo di quello disponibile alle fluttuazioni della materia ordinaria. L'ampiezza iniziale delle fluttuazioni di neutrini necessarie per spiegare le disomogeneità attuali della materia può essere, quindi, molto minore di quella richiesta nel caso di una miscela di radiazione e di materia normale. Con neutrini dotati di massa la variazione nella temperatura del fondo di radiazione necessaria per generare gli aggregati visibili di materia diminuisce di un ordine di grandezza o più, riconciliando così la teoria e le osservazioni.

La nuova versione della teoria delle frittelle conduce a una spiegazione naturale dell'origine della materia oscura nell'universo. Il collasso iniziale di una frittella distribuisce quasi tutti i neutrini su zone molto estese, perché la maggior parte di essi viene accelerata a grandi velocità, dell'ordine dei 1000 chilometri al secondo. Questi neutrini andranno a riempire le regioni oscure dello spazio intergalattico. Altri neutrini, però, si muovono più lentamente, perché inizialmente si trovano più vicini al piano centrale della frittella e non subiscono grandi accelerazioni. Il sottile strato di gas nelle vicinanze del piano centrale condensa e si suddivide per formare le protogalassie. I neutrini lenti vengono raccolti dagli aggregati di materia normale, e la materia al centro della protogalassia continua a condensare e forma stelle. I neutrini alla periferia della protogalassia, però, sono soggetti anche all'attrazione gravitazionale del resto della frittella e non condensano, diventando così la materia oscura dell'alone galattico.

Oggi è in corso di elaborazione una teoria più dettagliata della formazione delle galassie nel contesto della nuova teoria delle frittelle. Durante il collasso di una frittella, la componente neutrinica del gas che collassa attraversa il piano centrale della frittella senza interagire. La distribuzione di densità dei neutrini acquista discontinuità molto marcate, alcune delle quali si possono identificare con ammassi di galassie ricchi. Vladimir I. Arnold, un matematico dell'Università statale di Mosca, negli ultimi tempi ha collaborato con gli astrofisici a questo proposito, identificando questo tipo di discontinuità nella distribuzione complessiva della densità con alcune strutture elementari della branca della matematica chiamata teoria delle catastrofi.

La teoria delle frittelle, nella sua versione modificata, offre una profonda comprensione del carattere e dell'origine dell'attuale struttura dell'universo. Si basa su principi fisici ben noti e su ipotesi plausibili relative alle condizioni dell'universo nelle prime fasi della sua evoluzione. Come teoria dell'origine di una struttura su grande scala non è affatto unica, anche se oggi sembra che tanto la teoria quanto l'osservazione si muovano complessivamente nella direzione che abbiamo schematizzato. Nonostante ciò vi sono molti problemi importanti da risolvere prima che questa teoria si possa considerare stabilmente fondata.

Una volta ottenuta la conferma della teoria, se ne dovranno perseguire due linee di sviluppo. Innanzitutto sarà necessario rivolgere l'attenzione dei ricercatori alla struttura più fine dell'universo, alla formazione della prima generazione di stelle da un gas primordiale completamente privo di elementi pesanti. In secondo luogo bisognerà chiedersi come le condizioni iniziali ipotizzate dalla teoria delle frittelle si siano instaurate in epoche ancora precedenti della storia dell'universo. Oggi sono in corso diversi tentativi di far vedere come le fluttuazioni di piccola ampiezza richieste dalla teoria delle frittelle possano derivare da fenomeni precedenti molto meno noti. Questi tentativi si basano su teorie non ancora stabilite, ma i risultati preliminari lasciano spazio a un certo ottimismo riguardo alla possibilità di avere a disposizione una teoria coerente dell'universo entro la fine del secolo.

Il Mar Morto

Nel 1978 il lago più salato del mondo si è rimescolato completamente, disperdendo così un denso corpo di acqua salmastra fossile che le acque più dolci sovrastanti avevano isolato da ogni contatto con l'atmosfera

di Ilana Steinhorn e Joel R. Gat

Per migliaia di anni l'acqua dolce, con il suo leggero carico di sali asportati da rocce, sabbie e suoli, è scorsa verso sud attraverso la *rift valley* del Giordano fino al Mar Morto, un lago terminale; non ci sono, infatti, emissari. Sotto il cielo terso l'acqua che affluisce evapora nell'arida aria desertica, depositando i sali che si accumulano nel lago e nei suoi sedimenti. Per questo motivo il Mar Morto è il lago naturale più salato del mondo, con una salinità media di 280 grammi per chilogrammo, contro i 35 grammi che rappresentano la salinità media delle acque oceaniche. Il Mar Morto è anche il bacino più basso esistente sulla Terra; infatti la sua superficie si trova circa 400 metri al di sotto del livello del mare. Sia la salinità sia la particolare posizione topografica rendono il lago una notevole risorsa naturale; è già stato esplorato come «magazzino» di prodotti chimici da Israele e Giordania, e Israele ha progetti per utilizzare il potenziale del lago come salto per la generazione di energia elettrica, convogliando in esso acqua marina dal Mediterraneo.

Qualsiasi lago è un sensibile segnalatore di cambiamenti ambientali, le cui tracce sono impresse nei sedimenti del fondo. Nel caso del Mar Morto, però, la storia geologica e idrologica non è registrata solo nei sedimenti, ma anche nell'acqua salmastra stessa e nella sua struttura salina sul lungo termine. Tra tutti i laghi ipersalini del mondo il Mar Morto è il più salato, ma è anche il più profondo, con un massimo di profondità di 320 metri. Per secoli l'acqua dolce, scorrendo superficialmente, si mescolava solo in piccola parte con l'acqua molto più salata del lago tendendo così a formare strati meno salati, «galleggianti» sopra una densa colonna d'acqua fossile. In seguito, a partire dagli anni sessanta, la deviazione di quantità crescenti di acqua dolce per l'irrigazione provocò un abbassamento del livello del lago rendendone la superficie più salata e, quindi, più densa. Qualche anno fa l'antico gradiente scomparve e la colonna d'acqua si capovolse cosicché l'acqua salmastra fossile si mescolò con l'acqua più recente. In questo articolo descriveremo in breve la storia geologica, idrografica e chimica del Mar Morto, prima di esporre in particolare ciò che si sa a proposito di questo rimescolamento storico.

Il Mar Morto ha due bacini, uno a nord, più esteso e molto profondo, e uno a sud, più ridotto e profondo solo pochi metri. I due bacini sono separati da el-Lisan, in arabo «la lingua», una estesa penisola che si protende dalla sponda orientale formando un piccolo stretto lungo la sponda occidentale. Il lago è attualmente così poco profondo che la soglia dello stretto è emersa e il bacino meridionale è asciutto per la maggior parte dell'anno (a eccezione delle parti sfruttate come stagni di evaporazione). L'estremità meridionale del lago è all'incirca alla latitudine di transizione da un clima mediterraneo a un clima arido; i Monti della Giudea che lo limitano a occidente e i Monti di Moab a est hanno una precipitazione media annua di 600 millimetri.

Eppure il clima locale del Mar Morto è estremamente caldo e arido; il lago si trova infatti nel versante riparato dei Monti della Giudea e la precipitazione media annua è attualmente compresa tra 50 e 75 millimetri soltanto. A Sodoma, nel sud, per 300 giorni all'anno il cielo è terso, l'umidità relativa durante l'estate è compresa tra il 30 e il 40 per cento e la temperatura media mensile varia da 16 a 34 gradi centigradi; durante il giorno, in estate, si sono registrate anche temperature di 56 gradi centigradi. Considerando la grande aridità, l'acqua evapora dal Mar Morto con relativa lentezza, perché la tensione di vapore sopra la superficie è abbassata dai sali disciolti nell'acqua.

Negli anni trenta (prima che fossero iniziate le grandi opere di irrigazione), il fiume Giordano, che nasce dalla confluenza dell'emissario del Lago di Tiberiade (detto anche Mare di Galilea) a nord e di quello del bacino di drenaggio del fiume Yarmuk a est, forniva circa i due terzi dell'afflusso totale del Mar Morto. Il bilancio veniva completato da corsi d'acqua minori come l'Arnon da alcune sorgenti d'acqua dolce e d'acqua salata e dagli uidiàn stagionali che scorrono lungo i versanti montuosi. Tipico in una zona di transizione climatica, le fluttuazioni nelle precipitazioni e nel deflusso delle acque erano notevoli e la profondità e l'area del Mar Morto variavano di conseguenza.

Nel corso degli ultimi due decenni, si è assistito a una continua diminuzione dell'afflusso di acqua nel Mar Morto poiché, a partire dagli anni sessanta, gran parte dell'acqua dolce è stata deviata dapprima nel Sistema idrico nazionale di Israele e, in seguito, anche nel canale di el-Ghor, lungo il Giordano. Attualmente il livello del lago si abbassa di circa mezzo metro all'anno (a eccezione di occasionali periodi particolarmente umidi); a ogni modo, anche se si mantenesse l'attuale tasso di perdita per evaporazione, sarebbero necessarie parecchie centinaia d'anni prima che il lago si prosciugasse, dato che il bacino settentrionale è molto profondo. In realtà il tasso di perdita diminuirebbe dato che l'area superficiale verrebbe a ridursi in seguito all'abbassamento del livello e la salinità crescerebbe al diminuire del volume, abbassando così ulteriormente la tensione di vapore e quindi il tasso di evaporazione per area unitaria; alla fine, l'acqua salmastra sempre più concentrata diventerebbe igroscopica e assorbirebbe acqua dall'aria.

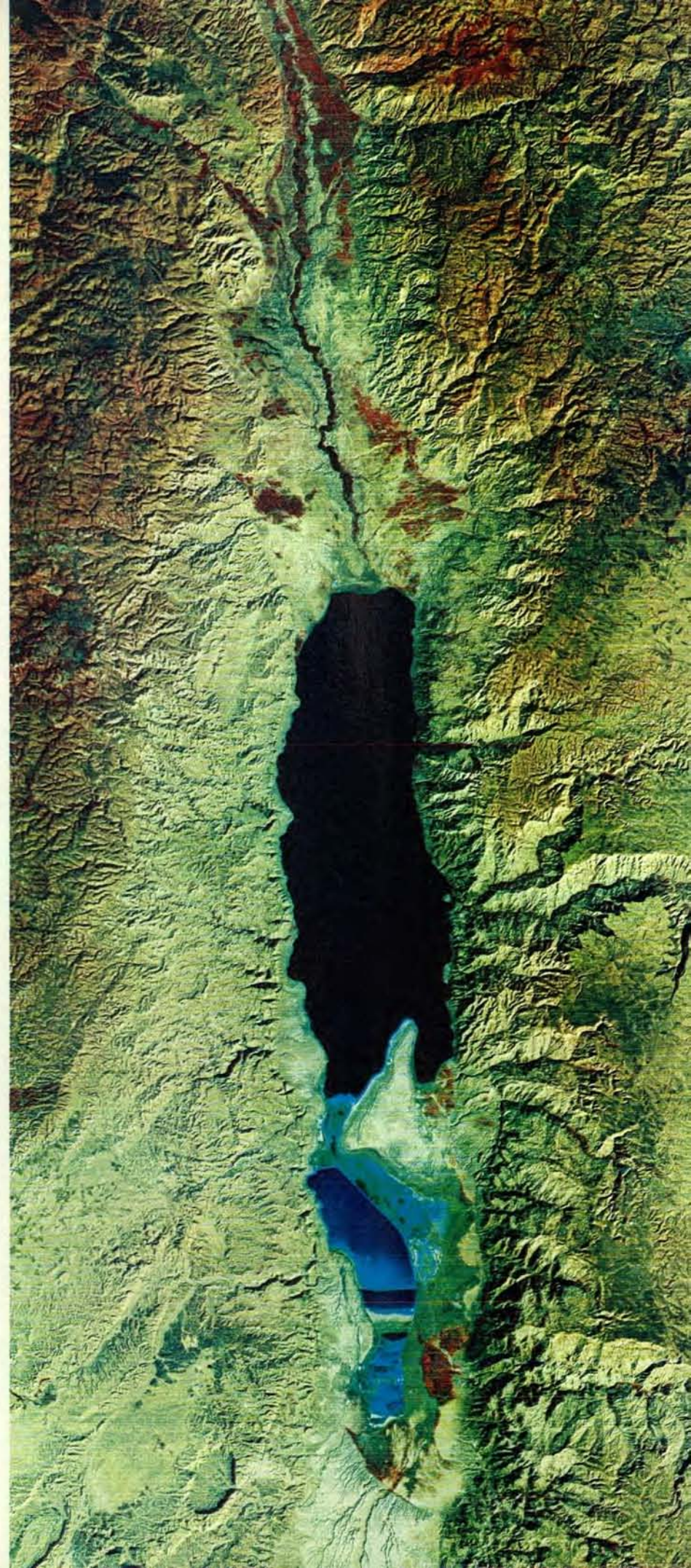
L'origine del Mar Morto può essere descritta in termini di tettonica a zolle: il movimento delle grandi zolle in cui è suddivisa la litosfera, ossia l'involucro rigido più esterno della Terra. Il lago giace sul Rift siro-africano, un margine trasforme lungo il quale le due zolle scorrono l'una contro l'altra da circa 10 milioni di anni: la zolla a est del rift si è mossa verso nord rispetto a quella a ovest. Il movimento relativo ha stirato e quindi assottigliato la crosta, che è andata incontro a subsidenza. Localmente la zona di faglia si sposta a zig-zag verso ovest trasversalmente alla direzione dello scorrimento che ha provocato l'apertura di bacini separati come quello del Mar Morto.

Immensi strati di sale (cloruro di sodio) e di gesso sul fondo della rift valley del Giordano rivelano che durante il Pliocene (da due a cinque milioni di anni fa, circa), la valle era in realtà una grande laguna, periodicamente collegata all'oceano verso ovest; il collegamento fu interrotto da movimenti tettonici, uno o due milioni di anni fa, e da allora la valle è stata occupata da una successione di laghi. Tra questi il meglio documentato, il lago Lisan, era più esteso del Mar Morto, molto più profondo (il suo specchio si trovava a 180 metri circa soltanto sotto il livello del mare) e anche meno salato. Forse 10 000 anni fa la valle e il Mar Morto assunsero quello che essenzialmente è il loro aspetto odierno, ma il livello del lago continuò a fluttuare leggermente.

Alcune allusioni contenute nella Bibbia suggeriscono che il livello dell'acqua doveva essere piuttosto basso, tanto che il bacino meridionale era asciutto; la più antica raffigurazione del Mar Morto, un mosaico del 560 d.C., mostra solo il bacino settentrionale. Una ricostruzione abbastanza attendibile delle oscillazioni del livello dell'acqua è possibile a partire dall'inizio del XIX secolo sulla base dei segni lasciati dalle inondazioni, sulla datazione del legname galleggiante o dei tronchi d'albero sommersi e su resoconti di viaggiatori (che riportano, per esempio, se il guado di Lisan, 402 metri sotto il livello del mare, poteva essere attraversato oppure no). Misurazioni reali, registrate fin dai tardi anni venti, riportano un livello massimo di 392 metri sotto il livello del mare nel 1930; da allora si è avuta una graduale diminuzione.

Le caratteristiche chimiche dell'acqua salmastra del Mar Morto si sono evolute insieme al lago e al suo bacino. L'acqua salmastra è costituita da una associazione tipica di sali. In confronto alle acque salmastre di origine marina, è ricca di calcio, magnesio, potassio e bromo, mentre è povera di sodio, solfati e carbonati. Una tale composizione non potrebbe essere il risultato dell'evaporazione di acque marine né della semplice alterazione dovuta agli agenti atmosferici di rocce continentali, né del passaggio in soluzione di minerali formatisi in precedenza per evaporazione. Gli stadi evolutivi dell'acqua salmastra del Mar Morto sono ancora controversi, ma deve esserci stata più di una origine.

Gli sforzi per determinare l'età dell'acqua salmastra portano a una conclusione analoga. Già nel 1715 l'astronomo inglese Edmund Halley suggerì che un confronto tra il tasso di apporto dei sali e un «inventario» chimico dei sali contenuti nel lago e nei suoi sedimenti avrebbe offerto la possibilità di apprendere in quanto tempo si

La Valle del Giordano e il Mar Morto sono visti da un'altezza di 570 chilometri in questa immagine Landsat a falsi colori. Le profonde acque salate del grande bacino settentrionale del lago assorbono luce e non la riflettono cosicché esso appare nero. Le piane melmose del bacino meridionale, più piccolo e prosciugato, appaiono in verde scuro e in azzurro; gli stagni di evaporazione cintati da argini presentano varie sfumature di blu, mentre la vegetazione appare rossa. L'immagine è stata ripresa nel dicembre del 1978, circa nel periodo in cui le acque del Mar Morto si stavano rimescolando.

era accumulata l'acqua salmastra (e così stabilire - secondo Halley - la data della Creazione). Nel 1961 Yaacov K. Bentor dell'Università ebraica eseguì questo tipo di calcolo con dati relativi al carico chimico del Giordano e di sorgenti e infiltrazioni salate note, giungendo ad assegnare all'associazione di sali del Mar Morto un'età approssimativa di 12 000 anni. Questo risultato implica che i sali contenuti nel lago siano quasi coevi al lago stesso, e che di conseguenza l'intero carico di sali depositati nei laghi che precedettero il Mar Morto siano sepolti al di sotto dei suoi sedimenti.

A un esame più attento, comunque, si scopre che la maggior parte dei sali in realtà non sono portati al lago dal sistema fluviale principale, bensì da sorgenti locali e da infiltrazioni. Tra queste ultime fonti molte hanno una composizione chimica alquanto vicina a quella del lago stesso: stanno riciclando infatti l'acqua salmastra del Mar Morto o quella dei suoi «predecessori». L'accumulo dei sali nel lago è quindi il risultato di un tasso d'apporto minore di quello calcolato da Bentor, ma per un arco di tempo maggiore. La particolare composizione dell'acqua salmastra rispecchia, quindi, una duplice origine: l'apporto recente e attuale dei prodotti dell'erosione continentale e il riciclaggio di depositi precedenti, inclusi quelli dell'antica laguna marina. Tale spiegazione è coerente con il fatto che il Mar Morto si trova nel punto più basso della superficie terrestre, dal quale l'asportazione dei sali accumulati è impossibile.

Le acque ipersaline non sono favorevoli alla vita, e fu infatti l'assenza di forme di vita superiori a dare il nome al Mar Morto. Benjamin Elazari-Volcani, dell'Istituto Sieff di Israele, potè stabilire nel 1936 che nel lago vi erano microrganismi. Egli trovò parecchie specie di batteri e una specie di alga verde, una *Dunaliella* che si trova comunemente in acque dolci o marine, ma può adattarsi anche ad ambienti ipersalini, sintetizzando grandi quantità di glicerolo viscoso il quale aumentando la pressione osmotica entro la cellula impedisce l'ingresso del sale.

Esiste una forte correlazione tra le proprietà fisiche e chimiche del Mar Morto e la quantità e i tipi di microrganismi che vi vivono. Uno studio compiuto in un periodo in cui il lago era stratificato ha rivelato che la quantità totale di batteri a 50 metri di profondità diminuiva di un fattore 100, mentre sotto i 100 metri non si trovavano alghe viventi. Nel 1970 alcuni ricercatori contarono, in superficie, 40 000 cellule di *Dunaliella* per millilitro; sotto i 50 metri trovarono solo batteri anaerobici.

Tra i minerali di importanza commerciale contenuti nell'acqua salmastra del Mar Morto vi sono il cloruro di potassio (KCl), che costituisce l'1 per cento dell'acqua salmastra; il cloruro di sodio (NaCl), l'8 per cento; il cloruro di magnesio ($MgCl_2$), il 13 per cento; lo ione bromo (Br^-), il 3 per cento. Il primo passo verso il loro sfruttamento commerciale fu compiuto negli anni trenta da M. A. Novomeyski, un ingegnere minerario della Siberia, che ottenne una concessione per l'estrazione di cloruro di potassio dall'amministrazione mandamentale britannica. Ancora oggi il processo di precipitazione selettiva di Novomeyski è alla base della produzione della Dead Sea Potash Works. L'acqua del lago viene trasferita in una serie di vasche di evaporazione poco profonde, dove ognuno dei vari sali precipita quando la sua concentrazione raggiunge il punto di saturazione. Il gesso (solfato di calcio) cristallizza per primo, seguito dal cloruro di sodio; la carnallite (un cloruro idrato di potassio e magnesio) inizia a cristallizzare quando l'acqua salmastra raggiunge un peso specifico di 1,3 grammi per centimetro cubo. Per estrarre il cloruro di potassio, impiegato soprattutto come fertilizzante, una fanghiglia di carnallite viene separata nei suoi componenti e il cloruro di potassio viene ulteriormente raffinato. Una piccola frazione della «salamoia» restante viene trattata con cloro gassoso per estrarre bromo.

In un primo tempo gli stagni di evaporazione erano sulla sponda, all'estremità meridionale del lago, ma la loro superficie totale fu estesa, nel 1966, fino a circa 130 chilometri quadrati per mezzo di sbarramenti costruiti nella metà del bacino meridionale poco profondo. Da quel momento, però, avvennero cambiamenti nel lago che influirono sull'impresa. L'aumento della salinità, infatti, abbrevia il tempo necessario per l'evaporazione, ma allo stesso tempo aumenta la precipitazione dei sali nelle tubazioni e nelle pompe. L'abbassamento del livello del lago inoltre implica che l'acqua salmastra deve essere pompata da zone più distanti e sollevata da profondità maggiori.

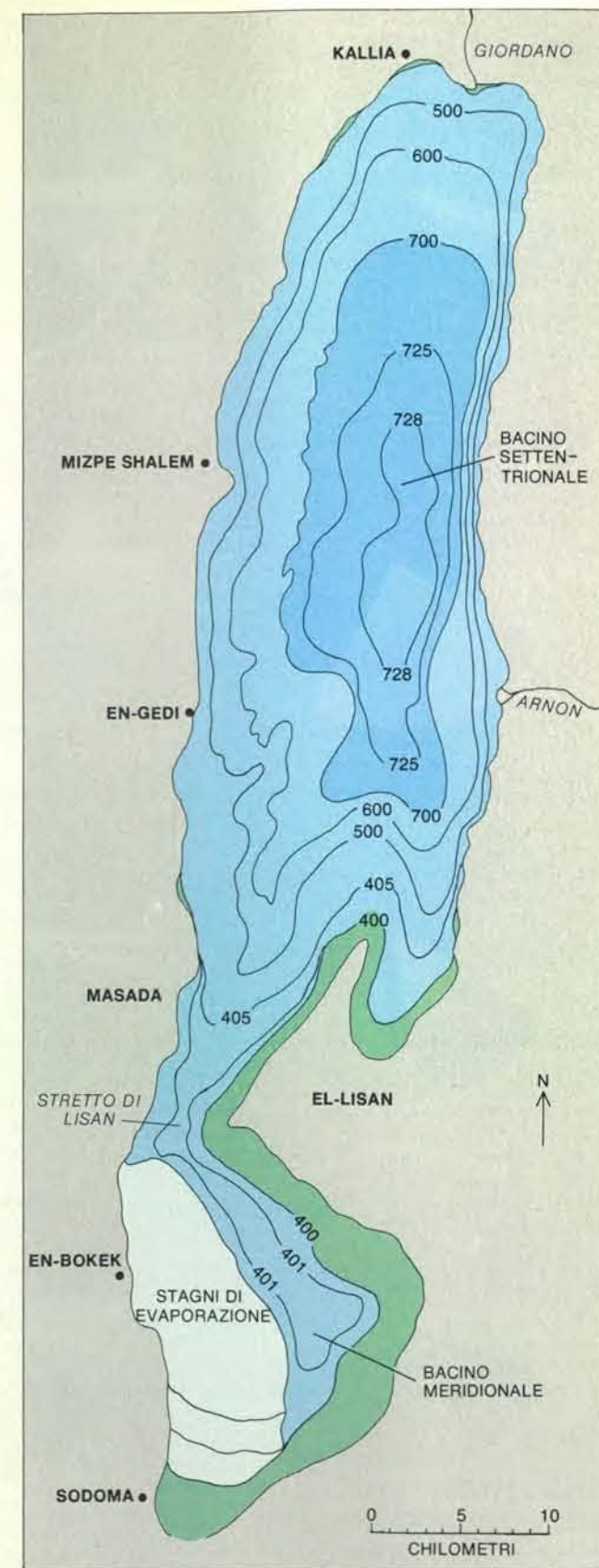
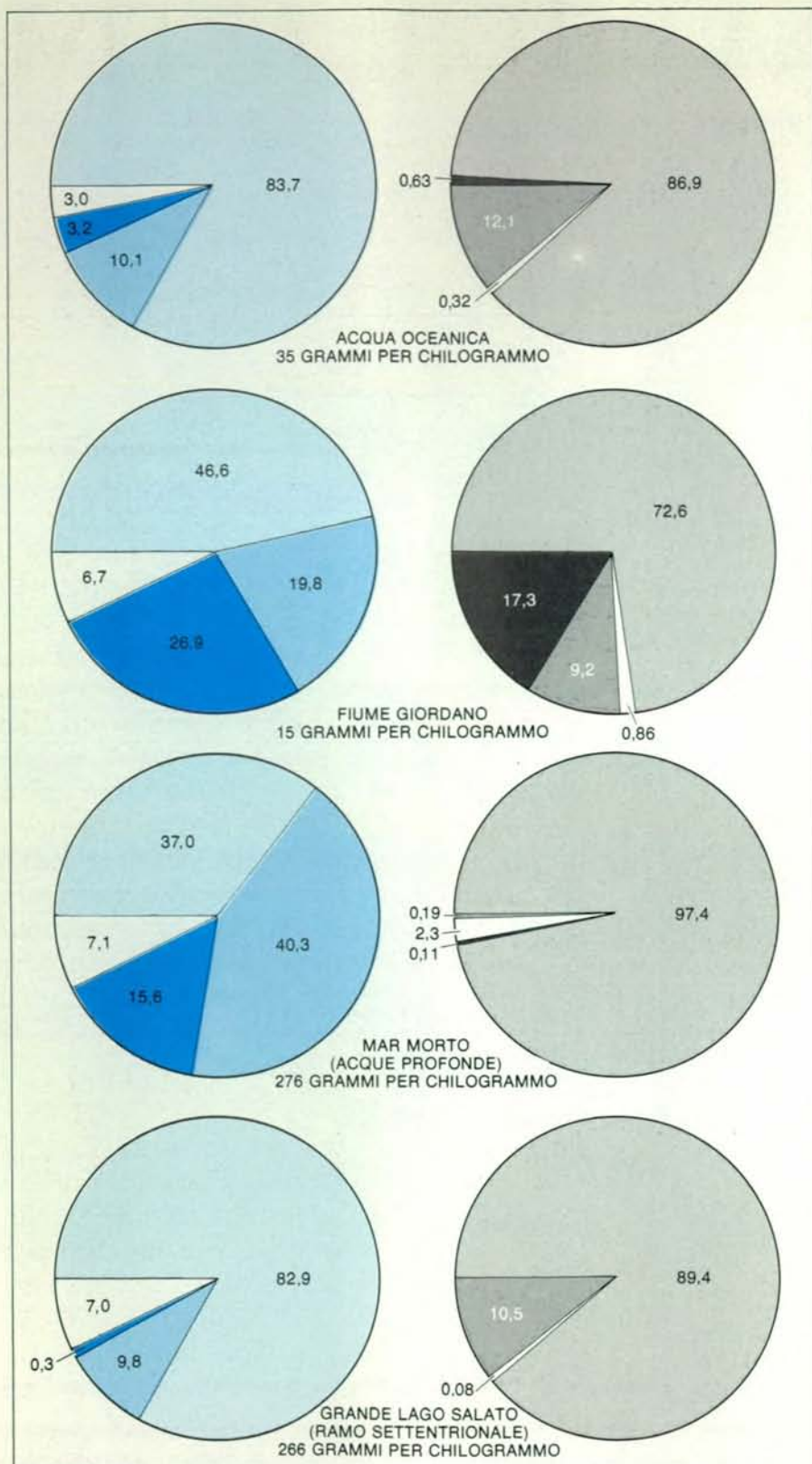
Nel 1902 Theodor Herzl, l'utopista della iniziativa sionista in Palestina, nel suo libro *Altneuland* suggerì che sarebbe stato possibile sfruttare i 400 metri di differenza tra il livello del Mediterraneo e quello del Mar Morto per la generazione di energia idroelettrica. Un vasto piano per la Valle del Giordano, avanzato 40 anni fa dal conservazionista americano Walter C. Lowdermilk, includeva la proposta di deviare l'acqua del Giordano per l'irrigazione e di costruire un canale che portando acqua dal Mediterrano compensasse la perdita di afflusso al Mar Morto e producesse 100 megawatt di energia elettrica. Attualmente il Giordano (così come lo Yarmuk) è stato in gran misura deviato cosicché l'abbassamento del livello del Mar Morto rende ora particolarmente attraente il progetto idroelettrico. È possibile infatti l'immissione di una grande quantità d'acqua senza dover sommergere i luoghi turistici e industriali che si trovano lungo le sponde israeliane del Mar Morto, almeno finché il lago non sia stato riportato al livello precedente agli anni sessanta.

Nel 1977, un gruppo di pianificazione incaricato dal governo israeliano esaminò quattro percorsi alternativi: uno dal Golfo di Aqaba, a sud, e tre dal Mediterraneo. Il percorso più settentrionale (quello considerato da Lowdermilk nella sua proposta) comprenderebbe un canale che andrebbe dai pressi di Haifa fino a sud del Lago di Tiberiade, attraverso la Valle di Jezreel, sommergendo la bassa Valle del Giordano: in pratica riprodurrebbe l'antico collegamento naturale tra il mare e la rift valley del Giordano. Questo percorso, però, sarebbe molto costoso e inciderebbe notevolmente sull'ambiente soprattutto su importanti falde acquifere. Gli altri tre percorsi richiederebbero numerose gallerie sotterranee. Il gruppo di studio fu favorevole al percorso più meridionale; non è il più breve, ma evita le principali falde acquifere del paese e potenzialmente può promuovere lo sviluppo del Negev settentrionale.

Il progetto richiede il riempimento del lago fino al livello tipico degli anni trenta, ossia fino a 393 metri al di sotto del livello del Mediterraneo, in un arco di tempo compreso tra 10 e 20 anni. Si potrebbe fornire una immissione annua di circa 1,6 chilometri cubi d'acqua che genererebbero circa 800 milioni di chilowattora di elettricità all'anno, consentendo di risparmiare qualcosa come 275 000 tonnellate di combustibile liquido. Verrebbero costruiti serbatoi di accumulo sulla ripida scarpata di Masada, a un'altezza di circa 300 metri al di sopra del lago, in modo che la generazione di elettricità possa essere regolata per soddisfare la domanda di picco.

Una volta che il Mar Morto sarà riempito, il tasso di immissione dovrà essere regolato in base al deficit medio del bilancio idrico. È stato calcolato che anche qualora il rimanente flusso del Giordano e quelli di locali corsi d'acqua fossero completamente deviati per l'irrigazione, il tasso medio di immissione d'acqua del Mediterraneo, che servirà a controbilanciare l'evaporazione, sarà soltanto circa due terzi del tasso di immissione previsto per la fase di riempimento cosicché il sistema idroelettrico diventerebbe inevitabilmente meno vantaggioso rispetto ai primi 10-20 anni. Il tasso di immissione consentito può essere aumentato nella misura in cui lo sia anche il tasso di evaporazione dalla superficie del lago. La tensione di vapore dell'acqua marina risulterà più elevata di quella dell'acqua salmastra del Mar Morto cosicché sarà vantaggioso evitare che l'acqua marina e quella del lago si mescolino in profondità, mantenendo di conseguenza uno strato nettamente distinto d'acqua marina mischiata a una parte di acqua salmastra alla sommità dell'acqua più densa del lago.

L'effetto della stratificazione è fondamentale anche nel nuovo sistema di produzione di energia, lo «stagno solare». Se sopra uno stagno di acqua salmastra viene mantenuto un sottile strato di acqua relativamente dolce, il Sole riscalda l'acqua salmastra la cui densità le impedisce di risalire e di disperdere il proprio calore nell'atmosfera. Il calore può essere estratto mediante uno scambiatore di calore che produce vapore, il quale a sua volta alimenta un generatore. In stagni solari sperimentali di piccola scala costruiti vicino al Mar Morto, sono state raggiunte temperature anche di 110 gradi centigra-

I bacini di drenaggio naturale del Mar Morto sono raffigurati in giallo e arancione. Gran parte dell'apporto del bacino idrografico settentrionale è stato deviato per l'irrigazione, riducendo in tal modo considerevolmente l'afflusso totale al lago. Il potenziale del lago come salto per la generazione di energia elettrica potrebbe essere sfruttato portandovi acqua dal Mediterraneo o dal Golfo di Aqaba. Sono qui illustrati i quattro percorsi proposti: attualmente è preferito quello che passa a sud di Bersabea. L'area il cui livello è inferiore a quello del Mediterraneo è indicata in rosa. Le linee tratteggiate indicano gli uidiàn stagionali.

I due bacini del Mar Morto, come risulta da questa carta batimetrica, sono molto diversi tra loro. Le profondità sono espresse in metri sotto il livello medio del mare. In azzurro e verde è illustrato il lago come era negli anni trenta, quando il livello della sua superficie era di 392 metri al di sotto del livello del mare. Nel 1975 la linea di costa corrispondeva circa alla isobata 400; nel bacino meridionale poco profondo, comunque, poteva ancora affluire acqua dal bacino settentrionale attraverso la soglia dello Stretto di Lisan. Attualmente il livello del lago è ancora più basso e il bacino meridionale è in gran parte prosciugato.

La composizione chimica particolare del Mar Morto viene confrontata con quella di altri corpi idrici. La salinità complessiva di ciascuno è data in grammi per chilogrammo. I diagrammi circolari mostrano la percentuale (in peso) calcolata per ogni ione positivo (cationi) e per ogni ione negativo (anioni). Appare chiaramente che l'associazione di sali del Mar Morto non potrebbe essere derivata semplicemente o dalle acque dell'oceano o dal Giordano.

di (una temperatura inferiore al punto di ebollizione dell'acqua salmastra). La produzione media di elettricità di un collettore solare di questo tipo potrebbe essere di cinque megawatt per chilometro quadrato di superficie del Mar Morto. Sono tuttora irrisolti numerosi problemi fisici e tecnologici sul galleggiamento degli stagni sulla superficie del Mar Morto e, di conseguenza, non è ancora stata stabilita la fattibilità del progetto.

Tutti i laghi che si trovano in un clima come quello del Mar Morto vanno incontro a un ciclo stagionale. Nel caso di un lago di acqua dolce, il Sole riscalda l'acqua superficiale durante l'estate diminuendone la densità, mentre l'acqua più profonda rimane più densa e più fredda. In base alla differenza di temperatura si crea un gradiente di densità che ha un effetto stabilizzante. L'acqua profonda indisturbata è isolata da un termoclino (una zona di confine in cui il gradiente è ripido) dai 10 metri d'acqua superficiali, riscaldati dal Sole e agitati dal vento. La circolazione completa, con il mescolamento delle acque superficiali e profonde, viene nuovamente stabilita in autunno o in inverno, quando l'acqua superficiale si raffredda a sufficienza. Nel Mar Morto la densità dell'acqua è influenzata sia dalla salinità sia dalla temperatura. Ne risulta un complicato schema di stratificazione, che può variare, ma non segue un semplice e ripetitivo ciclo annuale.

Le variazioni di salinità sono dovute o a diluizione dell'acqua salmastra per l'afflusso di acque dolci, o ad aumento della densità dell'acqua salmastra per evaporazione. In teoria, le variazioni di densità, causate da queste fluttuazioni nella salinità, potrebbero essere compensate da variazioni di temperatura nella giusta direzione; per esempio la densità rimarrebbe invariata qualora a un aumento di salinità di 9,2 grammi per chilogrammo corrispondesse un aumento di temperatura di 20 gradi centigradi. In realtà, però, durante questo secolo la salinità della superficie del Mar Morto ha subito variazioni comprese tra 40 e 50 grammi per chilogrammo, e variazioni ancora maggiori si ebbero in passato. Le grandi variazioni infatti, sia che risultino da un afflusso di acque dolci in eccesso rispetto all'evaporazione, o dalla situazione inversa, non possono essere compensate dal ciclo annuale di temperatura che dai 18 gradi centigradi circa in inverno raggiunge i 35 in estate. La variazione di temperatura dell'acqua non è in genere di entità sufficiente a controbilanciare quella di salinità. È difficile inoltre che i due tipi di variazione siano esattamente in fase.

Di conseguenza la struttura della colonna d'acqua del Mar Morto e la sua evoluzione stagionale e a lungo termine sono molto differenti da quelle di un lago d'acqua dolce e, data la grande profondità del Mar Morto, anche da quelle della maggior parte dei laghi salini e ipersalini. Le prime misurazioni idrografiche, compiute nel 1864, hanno rivelato che la colonna d'acqua del Mar Morto era stratificata in virtù

ACQUE A VARIABILITÀ STAGIONALE

ACQUE FOSSILI OMOGENEE

INVERNO

ESTATE

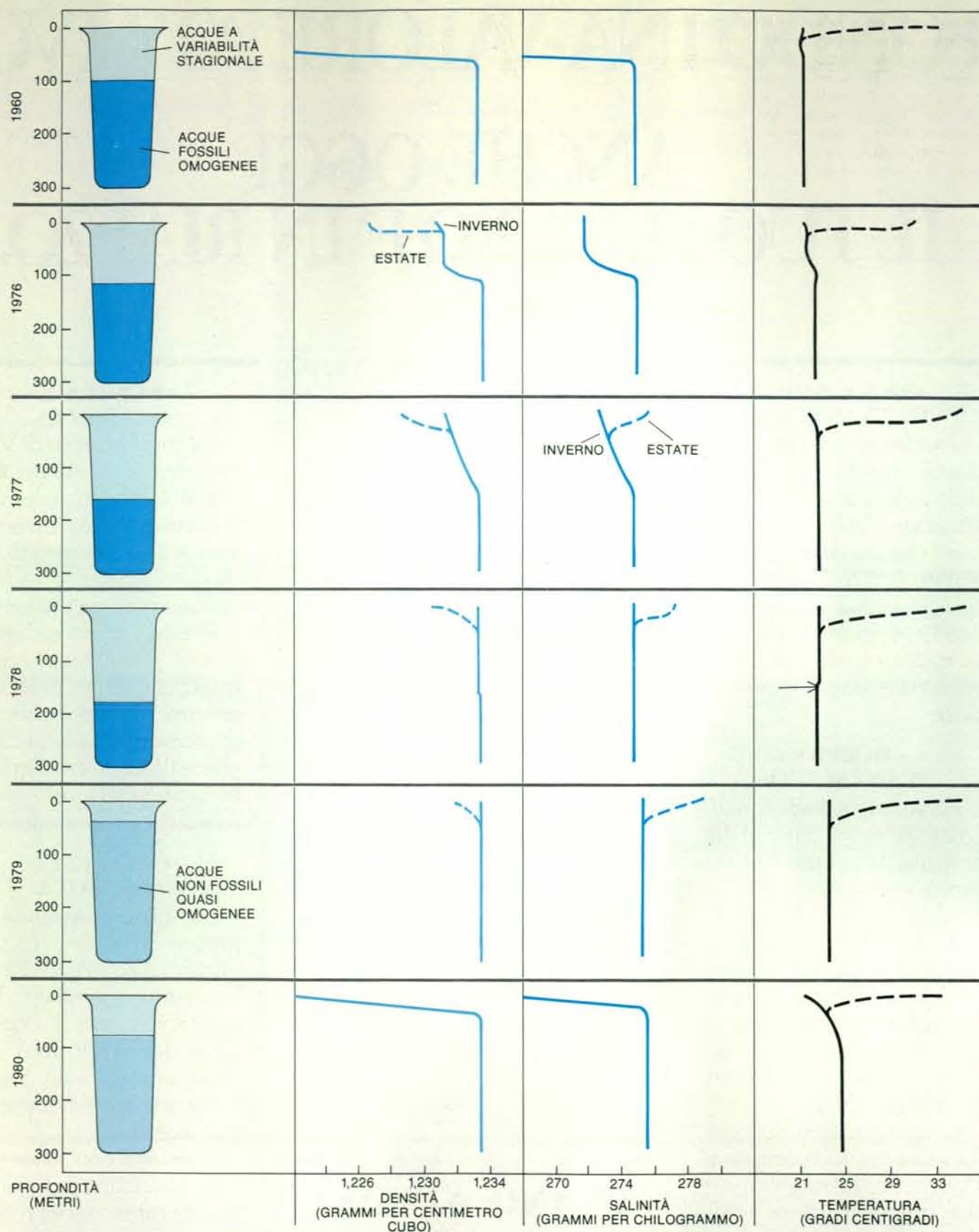
ACQUE NON FOSSILI QUASI OMOGENEE

1960 1976 1977 1978 1979 1980

PROFONDITÀ (METRI)

DENSITÀ (GRAMMI PER CENTIMETRO CUBO)

SALINITÀ (GRAMMI PER CHILOGRAMMO)

TEMPERATURA (GRADI CENTIGRADI)

Sono riportati in figura l'evoluzione della colonna d'acqua (*a sinistra*) e i profili di densità, di salinità e di temperatura che l'hanno influenzata dal 1960 al 1980. Nel 1960, come per molti anni precedenti, gli strati variabili superficiali erano separati dalle acque fossili profonde da una notevole differenza di densità (causata da una elevata variazione di salinità). Alla scala del grafico (alla quale i valori superficiali bassi escono dal campo rappresentato) i profili di densità e di salinità sono quasi corrispondenti. Nel 1976 l'interfaccia tra acque fossili e superficiali si era abbassata. La salinità e la densità degli strati superficali era aumentata, ma durante l'estate la densità superficiale era di nuovo diminuita a causa del riscaldamento. Queste tendenze continuarono nel 1977, ma durante l'estate si ebbe un'inversione del gradiente di salinità. Nel 1978 la salinità era uniforme in tutta la colonna; tuttavia una piccola differenza di temperatura (*freccia*) manteneva la stabilità del corpo d'acqua fossile. Nel 1979, dopo il rimescolamento globale, il corpo di acque fossili si era distribuito e l'intera colonna aveva una salinità e una temperatura più elevate. Nel 1980, dopo un inverno piovoso, si ebbe di nuovo un gradiente di salinità, ma vicino alla superficie.

di un gradiente di salinità. Questo fu confermato da misurazioni fatte nel 1919 e sembrava una caratteristica di lungo termine. David Neev del Servizio geologico israeliano e Kenneth O. Emery della Woods Hole Oceanographic Institution studiarono la stratificazione nel 1959 e nel 1960; essi calcolarono per gli strati superiori del lago una densità compresa tra 1,20 e 1,21 grammi per centimetro cubo (che corrisponde a una salinità media di circa 250 grammi per chilogrammo). Un gradiente di salinità - in quel caso una ben netta differenza di 25 grammi per chilogrammo a una profondità compresa tra 35 e 40 metri e una zona di transizione più graduale che raggiungeva gli 80 metri - separava questi strati superficiali, a variabilità stagionale, dalle acque più profonde. Entro gli strati superiori esistevano nette differenze di temperatura e di salinità, che variavano stagionalmente.

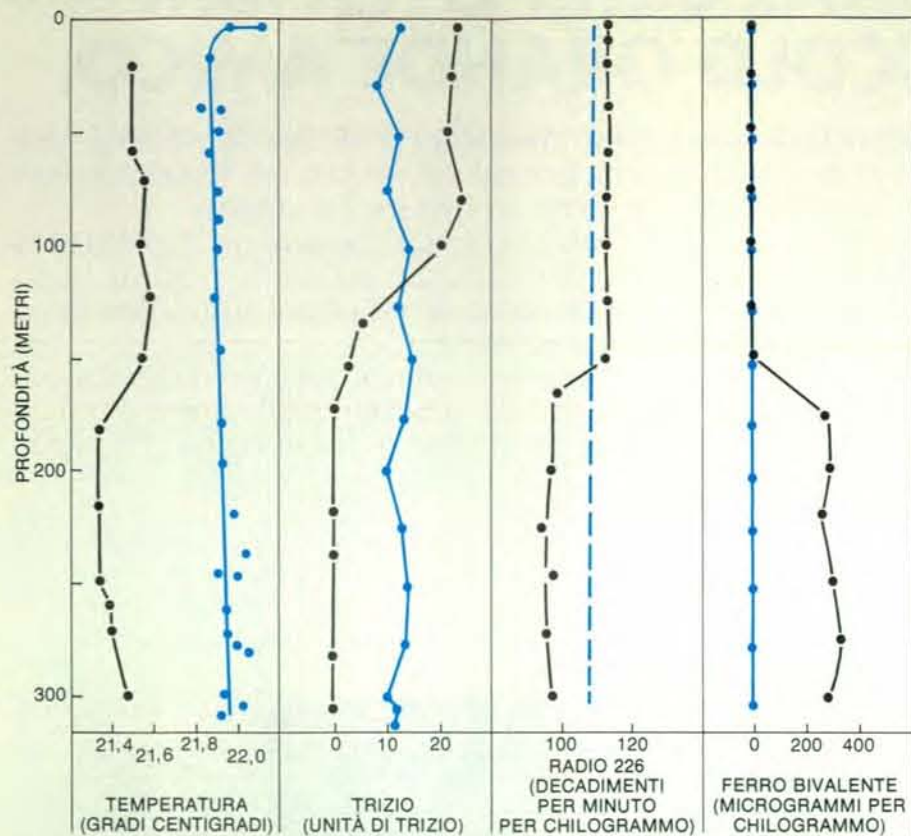
Sotto gli 80 metri di profondità l'acqua era essenzialmente omogenea, con una temperatura di 21,3 gradi centigradi e una salinità di 276 grammi per chilogrammo, un valore più elevato di tutti quelli registrati per campioni d'acqua superficiale nell'uno o nell'altro bacino. Inoltre, dai campioni di acqua profonda emanava un forte odore di acido solfidrico e, quindi, vi era ragione di pensare che l'acqua non contenesse ossigeno disciolto e ospitasse batteri anaerobici, che traggono energia dalla riduzione di solfati. Una conferma all'assenza di ossigeno si ebbe dal fatto che il ferro era presente nella forma ridotta, bivalente. Sembrava perciò che le acque omogenee al di sotto degli 80 metri costituissero un corpo di acqua fossile, rimasto per lungo tempo indisturbato, isolato dal contatto con gli strati superficiali e con l'atmosfera da un gradiente di densità.

Sono qui riportate la temperatura, la salinità e la densità dell'acqua a varie profondità in quattro fasi prima e dopo il rimescolamento del lago. Il grafico è un diagramma temperatura-salinità: una data temperatura e una data salinità insieme danno una particolare densità. Le diagonali sono isopicne, ossia curve di uguale densità (in grammi per centimetro cubo). Il breve profilo del gennaio 1978 mostra che le condizioni in superficie non erano molto diverse da quelle in profondità. In agosto le acque superficiali avevano raggiunto una salinità più elevata di quelle profonde, ma essendo più calde rimanevano ugualmente meno dense. Nel febbraio 1979, durante il rimescolamento, temperatura e salinità erano all'incirca uguali lungo tutta la colonna: il profilo è ridotto quindi a un unico punto. Nell'agosto 1980 un eccezionale afflusso di acque meno saline ristabilì temporaneamente un gradiente di densità e vi era un picnoclino poco profondo e brusco.

Il corpo di acqua fossile si disperse nell'inverno del 1978-1979 quando si mescolò con gli strati sovrastanti. Questo sensazionale rimescolamento e gli eventi che lo determinarono sono stati documentati da rilevamenti idrografici effettuati fin dal 1975 dal nostro gruppo del Weizmann Institute of Science, in Israele. Fin dall'inizio del lavoro fu chiaro che la salinità delle acque superficiali era aumentata in modo significativo: la differenza di salinità tra l'acqua profonda e l'acqua superficiale era passata da 25 grammi per chilogrammo nel 1959 a soli due grammi per chilogrammo nel 1975. La stabilità della colonna d'acqua era diminuita in proporzione. La sua omogeneità aumentava man mano che la salinità media degli strati superiori più recenti si avvicinava a quella dell'acqua fossile profonda. L'interfaccia tra le due masse d'acqua successivamente si abbassò: da una profondità compresa tra 60 e 80 metri nel 1960, scese a 80-100 metri nel 1975 e a 200 nell'autunno del 1978. Il corpo d'acqua fossile era dunque soggetto a «erosione», ma ciò che ne restava si conservava essenzialmente invariato.

Nell'estate del 1978 gli strati superiori erano effettivamente diventati più salati dell'acqua profonda; eppure la stratificazione si conservava ancora, e questa volta per la presenza di un gradiente termico. L'acqua superficiale era più calda e di conseguenza continuò per un certo periodo a essere leggermente meno densa rispetto all'acqua profonda. Durante l'inverno successivo l'acqua non fossile si raffreddò un poco e questo fu sufficiente per avviare il rimescolamento definitivo. Quando ciò avvenne, l'acqua superficiale era più salata e ancora un po' più calda dell'acqua fossile. Ne risultò che la colonna d'acqua resa omogena era, a quel tempo, di mezzo grado più calda e di solo quattro grammi per chilogrammo più salata di quanto non fosse stata l'acqua fossile.

Quale era stata la durata della stratificazione a lungo termine prima del rimescolamento? In altre parole, che età aveva il corpo d'acqua fossile? Marianna Stiller del Weizmann Institute e Y. C. Chung della Scripps Institution of Oceanography riuscirono a determinare l'età di due isotopi radioattivi trovati nel 1978 nella zona sottostante l'interfaccia dopo che si era abbassata: il radio 226 e il carbonio 14. Supposero che entrambi gli isotopi fossero stati introdotti negli strati superficiali e si fossero poi mescolati alla colonna d'acqua prima della sua stratificazione. (Il radio 226, infatti, ha origine soprattutto in sorgenti calde e salate lungo le sponde del Mar Morto; il carbonio 14, che è presente, in tracce, nell'anidride carbonica atmosferica, entra in soluzione nelle acque superficiali che sono a contatto con l'atmosfera.) Una volta che l'acqua scese in profondità e non ebbe più contatti con la superficie cessò l'immissione dei due isotopi che andarono incontro a decadimento radioattivo: il radio con un periodo di dimezzamento di 1260 anni e il carbonio di 5600 anni. Misurando l'entità del loro decadimento effettivo, si è potuto determinare che l'acqua al di sotto di 180 metri iniziò a essere isolata circa 300 anni fa, verso la fine della «piccola era glaciale» del XVIII secolo.

I calcoli descritti si basavano sulla supposizione che l'acqua profonda fossile costituisse un sistema assolutamente chiuso; in realtà nessun sistema di questo genere può essere completamente chiuso:

Il rimescolamento globale delle acque fossili con quelle superficiali è documentato da profili verticali di temperatura e di concentrazione di tre componenti chimici. Le misurazioni furono effettuate nel gennaio-febbraio del 1978 (*in nero*), un anno prima del rimescolamento, e nel febbraio successivo (*in colore*), durante il rimescolamento. La linea tratteggiata indica valori stimati. I profili antecedenti il rimescolamento mostrano una significativa differenza tra le caratteristiche dei due tipi di acqua; tali differenze non esistono più dopo il rimescolamento globale. Per il lungo isolamento l'acqua fossile aveva proprietà chimiche e fisiche particolari: valori bassi per il trizio e il radio e la presenza di ferro bivalente, che indica mancanza di ossigeno.

esistono sempre scambi per diffusione, anche attraverso un picnoclino, ossia una zona di confine tra densità diverse. In quale misura dunque l'acqua profonda fu realmente isolata durante la stratificazione a lungo termine?

Nel 1967 Abraham Lerman del Weizmann Institute calcolò quale tasso di rimescolamento attraverso il picnoclino ci si poteva aspettare semplicemente in base al moto browniano casuale delle molecole e concluse che il rimescolamento per diffusione molecolare sarebbe stato un processo molto lento e avrebbe richiesto migliaia di anni. In realtà il quadro non è così semplice. Sul fondo del lago vi sono sorgenti e infiltrazioni; talvolta un «pennacchio» di densa acqua salmastra superficiale del bacino meridionale può immergersi in profondità nel bacino settentrionale. I profili termici nella massa d'acqua profonda e nella zona di transizione permettono di calcolare il tasso di afflusso di acqua calda (e forse salata) nella massa d'acqua profonda. Il sistema, perciò, non è realmente sigillato. A ogni modo misurando traccianti radioattivi a vita breve come il trizio (che proviene dall'atmosfera), è possibile calcolare un limite superiore al tasso di rinnovamento dell'acqua fossile; indubbiamente deve essere un processo lento, la cui scala temporale è almeno di 1000 anni.

Anche negli anni immediatamente precedenti al rimescolamento (quando cioè la stabilità della stratificazione era diminuita, aumentando le probabilità che piccole porzioni di acqua superficiale potessero immettersi nella massa d'acqua profonda) la temperatura e i dati relativi al trizio indicano che ogni anno veniva rinnovata meno del 2 per cento dell'acqua fossile. Soltanto il rimescolamento finale avvenuto nell'inverno 1978-1979 permise la discesa di acqua dalla superficie e il suo mescolamento con il corpo d'acqua fossile.

Quale combinazione di condizioni può causare un rimescolamento di questo tipo nelle acque del Mar Morto? Indubbiamente il requisito essenziale è la diminuzione della stabilità della colonna d'acqua in seguito a un aumento della salinità delle acque superficiali, dovuto a una diminuzione del livello dell'acqua. Perché tutto questo possa accadere, negli strati superficiali, più recenti, devono esistere le condizioni favorevoli a un rimescolamento verticale. Tali condizioni non possono esistere nel Mar Morto durante l'estate, quando gli strati superficiali sono riscaldati e quindi si stabilisce un picnoclino controllato dalla temperatura. Inoltre durante l'inverno o la primavera un afflusso di acque dolci può diluire l'acqua salmastra superficiale e stabilire così un picnoclino basato su un gradiente di salinità. In genere è in autunno o nel primo inverno che un picnoclino, vicino alla superficie, scende in profondità a causa del raffreddamento degli strati superficiali; la profondità alla quale avviene il rimescolamento dipende dallo stato in cui in quel momento si trova la colonna d'acqua. Spesso, dunque, si incontra (andando verso il basso) una gerarchia di strati di età crescente che vanno da acque molto recenti nella zona superficiale fino allo strato fossile al fondo.

È solo dopo più stagioni secche che gli strati superficiali possono raggiungere un grado di salinità sufficiente (e quindi un grado di densità sufficiente) perché il mescolamento autunnale raggiunga lo strato di transizione che delimita il corpo d'acqua fossile, e in seguito il fondo del lago, causando così un rimescolamento totale. D'altra parte da una stagione particolarmente piovosa può derivare la persistenza di un picnoclino a piccola profondità per parecchi anni. Dopo la stagione 1979-1980, che fu molto umida, per esempio, la colonna d'acqua (che era stata resa omogenea durante il rimescolamento) subì una stratificazione che durò tre anni, fino alla fine del 1982. Questo periodo di tre anni, comunque, non fu sufficiente per ripristinare la natura «fossile» dell'acqua profonda.

Vi sono stati altri rimescolamenti in passato? Nella ricerca di dati che registrino tali eventi si potrebbe iniziare a considerare che una condizione necessaria per un rimescolamento totale (anche se forse non sufficiente, poiché il fenomeno può essere evitato da inondazioni occasionali) è un periodo di acqua relativamente bassa (e conseguente salinità media elevata). Durante il recente rimescolamento la salinità elevata provocò la precipitazione di salgemma (forma minerale del cloruro di sodio) dall'acqua salmastra del Mar Morto. Qualsiasi rimescolamento precedente potrebbe essere stato accompagnato da una precipitazione analoga, rivelabile dalla presenza di cristalli di salgemma nei sedimenti. Una ricerca accurata di strati di salgemma precipitati nella colonna sedimentaria potrebbe fornire la prova di periodi di siccità che potrebbero essere stati associati a precedenti rimescolamenti nella lunga storia del Mar Morto.

Come è giusto aspettarsi, le variazioni di salinità che abbiamo riscontrato ebbero una notevole influenza sul biota del Mar Morto. Nel 1979, dopo che anni di acque basse e di aumento della concentrazione salina avevano portato al rimescolamento, il lago sembrava privo di vita. Venne in seguito il piovoso inverno 1979-1980 e il numero di mircorganismi aumentò in modo impressionante, arrivando in superficie a 19 milioni di cellule per millilitro. La fioritura di microrganismi cambiò le proprietà ottiche dell'acqua superficiale: il lago passò dal blu verdastro al blu rossiccio, fornendo la prova che il Mar Morto non è affatto morto.

Sei millenni di caccia al bisonte

A partire dal 3700 a.C. fino all'arrivo dei bianchi, in un sito del Canada occidentale chiamato Head-Smashed-In, i cacciatori indiani uccidevano intere mandrie di bisonti spingendole giù da un precipizio

di B. O. K. Reeves

Secondo molti resoconti e illustrazioni dell'inizio dell'Ottocento, gli indiani delle Great Plains dell'America Settentrionale cacciavano i bisonti inseguendoli a cavallo. Ciò farebbe pensare che questi popoli non abbiano avuto molto successo nella caccia al bisonte se non dopo l'introduzione dei cavalli nell'America Settentrionale da parte degli europei, avvenuta nel Cinquecento. Scavi archeologici condotti nelle vaste praterie a est delle Montagne Rocciose contraddicono però questo assunto. I primi paleoindiani entrati nella regione circa 10 000 anni fa furono cacciatori temibili, anche se cacciavano a piedi. Dapprima i mammut furono la loro preda preferita e, quando questi grandi animali cominciarono a scarseggiare, essi si dedicarono alla caccia di due specie diverse di bisonte: il grande *Bison antiquus* dalle corna diritte e il più piccolo *Bison occidentalis,* oggi entrambi estinti. I successori dei paleoindiani delle pianure continuarono a cacciare i bisonti a piedi, nel corso di battute di caccia collettive.

Qui io riferirò su ciò che è stato riportato in luce in uno dei siti in cui veniva praticata questo tipo di caccia collettiva al bisonte, sito che si trova 130 chilometri a sud di Calgary, al margine sudorientale delle Porcupine Hills, nell'Alberta occidentale. Qui gli indiani piegan dei tempi storici e i loro predecessori hanno dato la caccia ai bisonti per almeno 5600 anni, e forse anche per 9000, spingendoli giù da un dirupo. Come nel caso di altri siti in Canada e negli Stati Uniti, in cui veniva praticato questo tipo di caccia, la topografia del luogo era favorevole all'avvicinamento furtivo dei cacciatori, che avevano buon gioco nel mettere una mandria in una fuga disordinata che si concludeva con un salto da un precipizio. I piegan chiamano il sito Head-Smashed-In, ossia «testa fracassata».

Il precipizio usato dai cacciatori è un affioramento di arenaria che è stato alterato dagli agenti atmosferici e che è crollato nel corso dei secoli, cosicché al suo piede si è formata una scarpata di detriti costituita da lastre staccate di arenaria e da frammenti minori. Oggi il salto verticale dal margine del precipizio alla sommità della scarpata di detriti misura fra nove e 10 metri. A ovest del precipizio le valli di sei affluenti di un fiume centrale, l'Olsen Creek, occupano un bacino poco profondo, di circa 36 chilometri quadrati di estensione, orlato da un altopiano continuo tranne che nel punto in cui l'Olsen Creek esce dal bacino. Più di 500 *cairn,* cumuli di pietre con un diametro di un metro circa e un'altezza di circa 30 centimetri, segnano ancora tratturi lunghi anche otto chilometri lungo i quali i cacciatori incalzavano i bufali sino al dirupo.

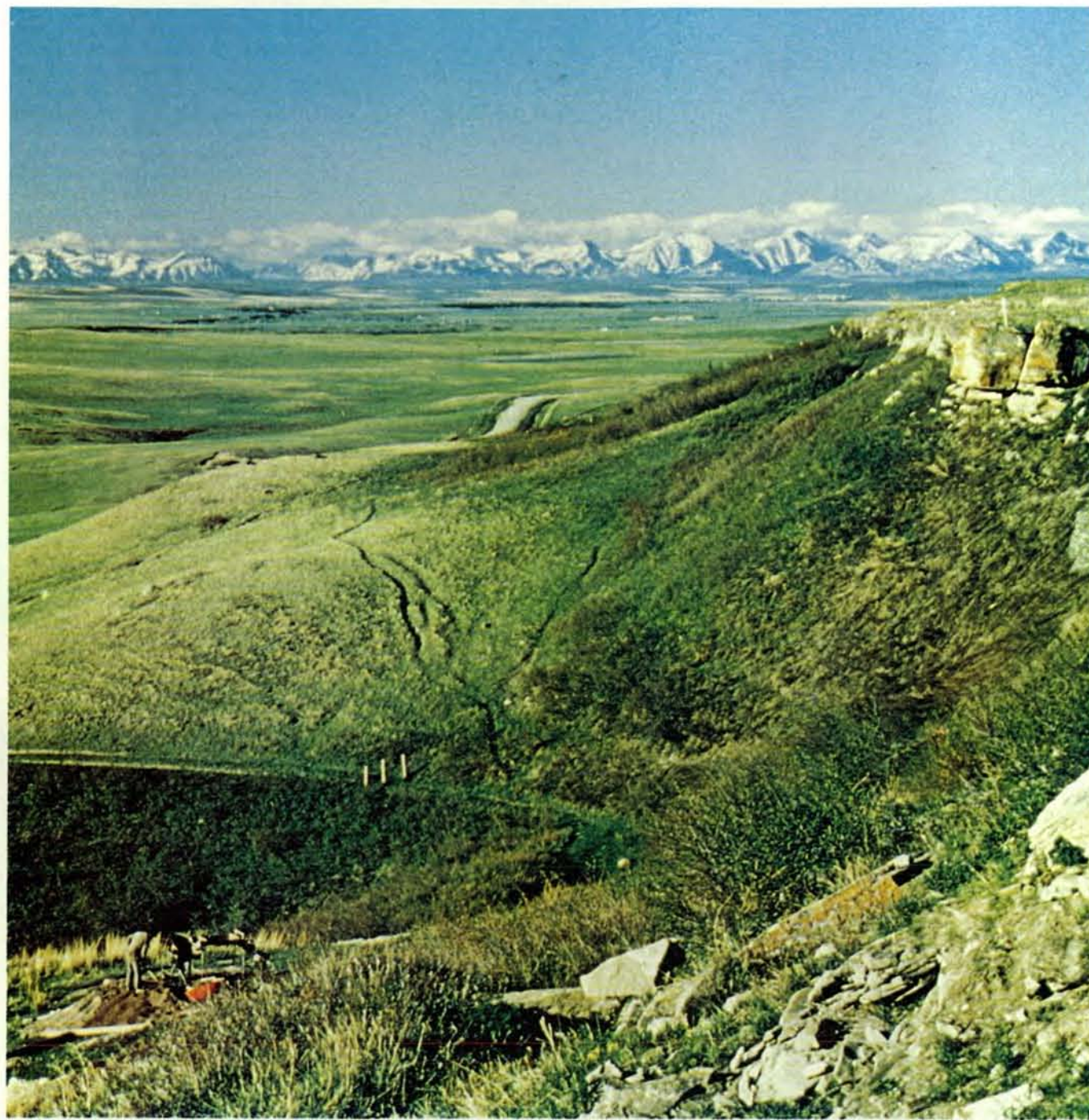
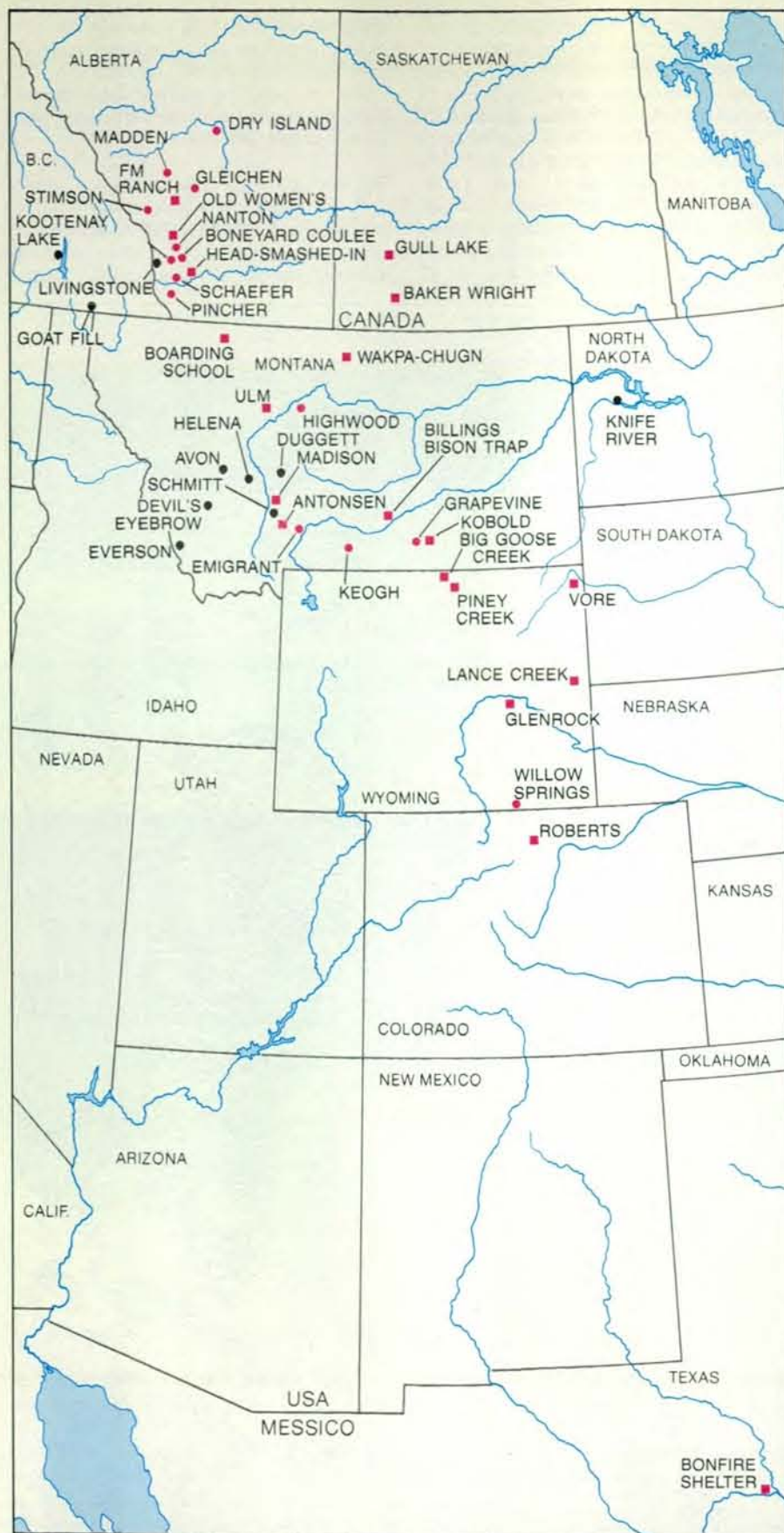
La scarpata di detriti al di sotto del precipizio non fu formata soltanto dall'erosione. Essa è in gran parte il risultato del distacco dalla faccia del dirupo e del successivo movimento verso il basso di due grandi blocchi di arenaria distinti. La prima di queste due cadute di blocchi avvenne attorno al 3700 a.C. e la seconda all'inizio dell'Era volgare. È difficile stimare quale fosse il dislivello fra il bordo del precipizio e la scarpata di detriti dopo la prima frana, ma anche dopo la seconda era quasi doppio rispetto al salto attuale.

L'uccisione di bisonti attuata facendoli cadere da un precipizio non fu il solo metodo di caccia praticato dagli indiani delle Great Plains del IV millennio a.C. A giudicare dai reperti archeologici e dalle tecniche usate dai piegan in tempi storici, i bisonti venivano sospinti talvolta in paludi, in corsi d'acqua o in gole senza uscita e persino in spazi chiusi o in recinti costruiti allo scopo dai cacciatori. Il primo bianco che visitò il paese dei piegan, un commerciante della Hudson's Bay Company di nome Peter Fidler, ci ha lasciato una narrazione molto viva di questo tipo di caccia al bisonte, cui egli ebbe modo di assistere a Stimson Creek nel dicembre 1797. «Il nostro capo è il Pound Master - scrisse Fidler - e ogni volta che una mandria [di bisonti] entra nello spazio recintato egli deve andare a uccidere il primo. Generalmente gli spara con un fucile. I giovani uccidono poi gli altri bisonti con frecce, baionette legate all'estremità di pertiche ecc. Spesso si usa anche l'accetta, ed è impressionante vedere quei poveri animali così rinchiusi senza alcuna possibilità di fuggire... Qualcuno, con un colpo di accetta, squarcia quasi completamente un fianco di un bisonte e il povero animale corre per qualche tempo, percorrendo un tratto considerevole per tutto lo spazio recintato, trascinandosi dietro le viscere, che vengono calpestate dagli altri animali.»

Nel corso di un soggiorno di sei settimane a Stimson Creek, Fidler stimò che i piegan avessero intrappolato e ucciso nello spazio recintato, che misurava 30 metri per 45, più di 250 bisonti. Avrebbero potuto catturarne di più, scrisse, se non fosse stato che «quando il vento soffiava dal recinto verso le tende, c'era un fetore intollerabile per il gran numero di carcasse pietrificate [*sic*], ecc. e fu questa la ragione che ci indusse ad andarcene».

Junius B. Bird, dell'American Museum of Natural History, fu il primo archeologo di professione a compiere scavi a Head-Smashed-In, ma i risultati delle sue ricerche, condotte nel 1938, non furono mai pubblicati. Nel 1949 Boyd Wettlaufer, che era allora studente all'Università del New Mexico, eseguì scavi di sondaggio sia nella scarpata di detriti sia nel sito di un accampamento circa 18 metri sotto il precipizio, dove venivano trasportate le parti principali dei bisonti uccisi per procedere a un sezionamento più minuto. La Provincia di Alberta eresse un *cairn* a Head-Smashed-In per commemorare le sue ricerche, che si pensava fossero le prime condotte nell'Alberta. Il *cairn* ebbe però l'effetto di attrarre scavatori illegali, il cui interesse fu ulteriormente sollecitato dalla notizia di uno scavo condotto nel 1958 nel sito vicino di «Old Women's», dove, in prossimità di un precipizio, i resti dell'attività di caccia si erano accumulati sino a raggiungere un'altezza superiore ai sette metri, e dove gli strati più antichi risalivano al I secolo d.C.

Vedendo che l'integrità del sito era in pericolo, il Glenbow-Alberta Institute e l'Università di Calgary unirono le forze per organizzare una importante iniziativa di recupero nel corso di due stagioni (1965-1966). Nei tre mesi della prima stagione, i miei collaboratori e io lavorammo principalmente sulla parte meridionale della scarpata di detriti, scavando un'area di tre metri per sei. A una profondità di otto metri l'area si ridusse a un pozzo di 1,5 metri quadrati, che continuammo a scavare sino a raggiungere la profondità di 11 metri. Al di sotto del livello di otto metri, il pozzo passò attraverso uno strato di terra sterile, prima di ritrovare, più in profondità, altre tracce di caccia al bisonte. Le dimensioni ridotte del pozzo ci impedirono di procurarci, a questo livello più basso, ossa di bisonte o punte di armi da lancio in quantità adeguate per accertare fasi di occupazione e date. Il materiale più abbondante trovato al livello di otto metri di profondità risaliva però al 1000 a.C. circa, e noi eravamo fiduciosi sul fatto che gli strati più bassi si sarebbero rivelati considerevolmente anteriori.

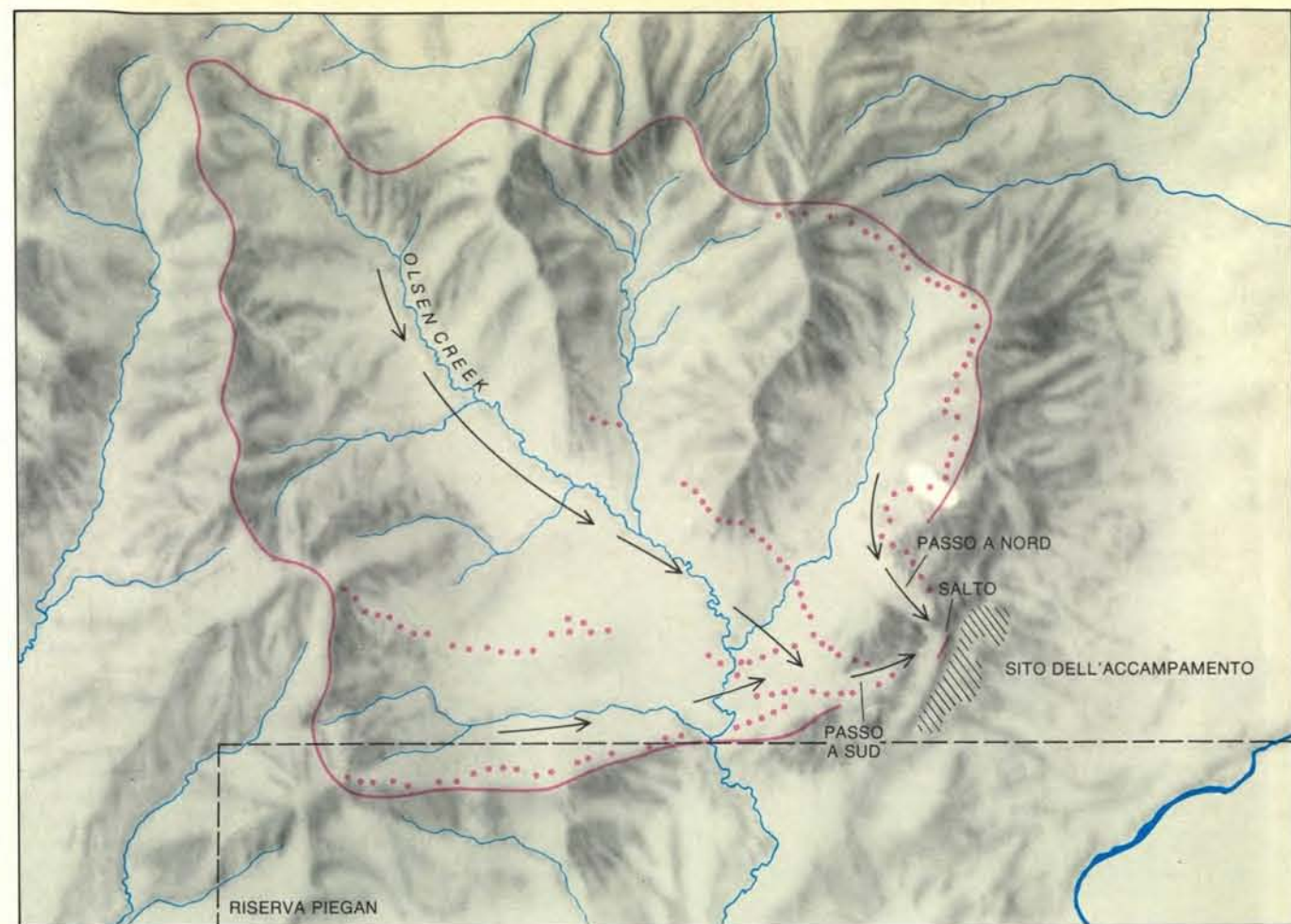
Nella seconda stagione i nostri sforzi principali si concentrarono sulla scarpata di detriti circa 100 metri a nord del pozzo del 1965. Una delle unità di scavo aperte qui, con una superficie di tre

La prateria si estende al centro di quest'immagine del margine sudorientale delle Porcupine Hills nella provincia di Alberta in Canada. Gli uomini che scavano, in primo piano a sinistra, sono membri del gruppo di ricercatori cui appartiene l'autore, i quali hanno studiato un sito in cui gli indiani delle praterie praticavano la caccia al bisonte facendo precipitare le mandrie da un dirupo. Questi cacciatori fabbricavano punte e coltelli in pietra ricavandoli da ciottoli di selce e quarzite estratti dai depositi morenici glaciali della prateria. Altri utensili erano fatti però di materie prime importate, come calcedonio proveniente dalle Montagne Rocciose, visibili qui all'orizzonte.

Nelle Great Plains dell'America Settentrionale gli archeologi hanno recentemente scoperto numerosi siti in cui fu praticata la caccia al bisonte. Questi siti si collocano in un intervallo di tempo che va dall'8000 a.C. circa, nei tempi paleoindiani, a periodi post-paleoindiani comprendenti l'arrivo dei coloni europei. I siti, scavati e non scavati, dove la caccia al bisonte veniva praticata ricorrendo al salto da un precipizio, tanto nel Canada quanto negli Stati Uniti, sono indicati rispettivamente con quadrati e pallini in colore. I luoghi d'origine dei tipi di pietra usati per produrre punte e altri utensili dagli occupanti del sito dell'Alberta, Head-Smashed-In, sono indicati con pallini neri.

metri per tre, fu scavata sino a una profondità di quattro metri prima di raggiungere blocchi di arenaria caduti. Lo strato più basso - uno strato di sabbia di colore grigio dorato, spesso da mezzo metro a un metro - conteneva un deposito di ossa di bisonte. L'analisi al carbonio 14 del collageno delle ossa indicò la data del 3710 a.C. con un'approssimazione in più o in meno di 100 anni. Alle ossa erano associate punte di armi da lancio con un incavo su ciascun lato, le quali, pur essendo fatte di pietra locale, sono dello stesso tipo di quelle trovate qualche anno fa in un sito del Wyoming noto come Mummy Cave, 850 chilometri più a sud.

Il blocco di arenaria più antico, situato inferiormente agli strati di ossa più vecchi sia nell'area più a nord sia in quella più a sud della scarpata di detriti, ha uno spessore di tre metri. I geologi del Governo canadese, che ci aiutarono nello scavo della stagione 1966, si spinsero in profondità fino a raggiungere suoli sepolti che erano stati trasportati via dall'acqua da sotto il blocco. In questi materiali essi trovarono semplici utensili ricavati da ciottoli e ossa di bisonti macellati. Il nostro gruppo non poté però sfruttare la loro scoperta a causa sia della profondità della fossa scavata dai geologi sia della sua instabilità. Il loro profondo scavo non fornì una quantità di ossa di bisonte sufficiente per eseguire l'analisi al radiocarbonio, cosicché l'età esatta di questi strati, che sono chiaramente i più bassi a Head-Smashed-In, rimane imprecisata.

Un possibile indizio circa la loro età proviene da una scoperta casuale fatta nel 1948 da Wettlaufer. Fra gli oggetti trovati in superficie, nelle sue ricerche condotte per una sola stagione, c'era una punta di arma da lancio paleoindiana. Essa era evidentemente inclusa nel suolo dragato durante la costruzione di un bacino d'acqua per il bestiame che viene ora allevato nell'area. Se la punta, appartenente inconfondibilmente alla tradizione paleoindiana nota come Cody Complex, era associata in origine ai sedimenti dilavati da sotto il blocco di arenaria più antico, Head-Smashed-In potrebbe essere stato un *buffalo jump* già nel VI millennio a.C.

In ordine di età decrescente, i resti culturali da noi riportati in luce nelle nostre due stagioni di scavi (e in un altro mese di lavoro compiuto nello stesso sito nel 1972) forniscono la documentazione di un uso costante, anche se non continuo, del salto dei bisonti almeno a partire dal 3700 a.C. sino all'inizio dell'Ottocento. Vorrei compendiare le prove archeologiche dell'uso del sito da parte di quattro successive culture preistoriche e del suo uso finale da parte di cacciatori piegan in tempi moderni. Per arricchire gli scarni dati della documentazione archeologica, aggiungerò qualche commento su ciò che i manufatti ci dicono circa reti di scambio e possibili spostamenti di popolazioni nello stesso periodo, a cominciare dalla presenza nel sito di membri della cultura della Mummy Cave fra il 3700 e il 3100 a.C.

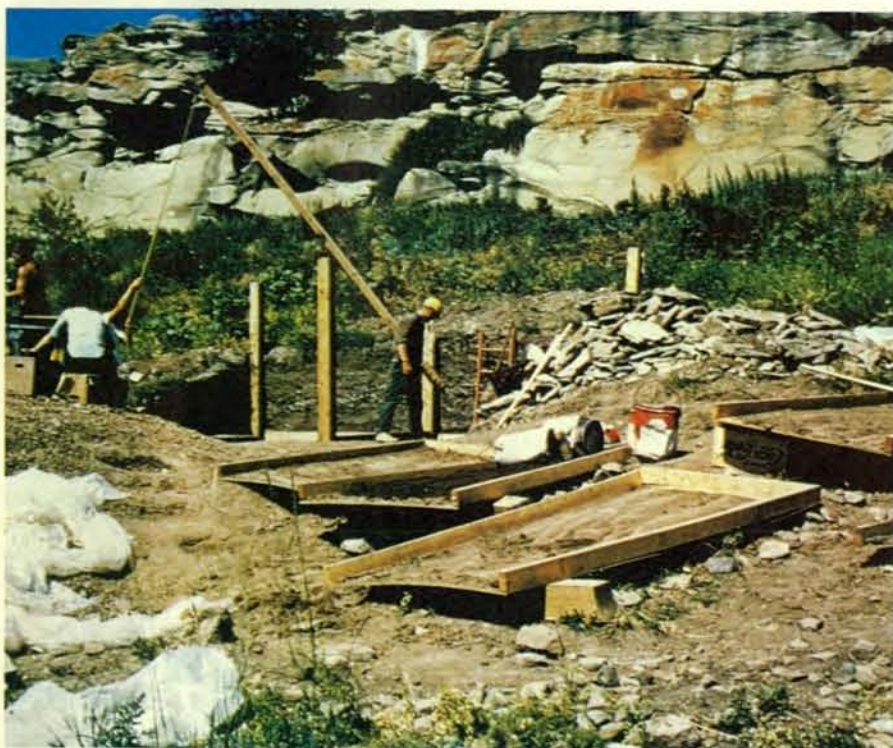
Due notazioni preliminari, una climatica e l'altra geologica, ci saranno d'aiuto

Il bacino dell'altipiano in cui gli indiani inseguivano i bisonti sino al precipizio di Head-Smashed-In copre un'area di 36 chilometri quadrati; nel bacino le valli dei sei affluenti dell'Olsen Creek forniscono tratturi naturali (*frecce*) che consentivano di spingere i bisonti verso il salto. Le file di cumuli di pietre (*pallini in colore*) aiutavano i cacciatori a guidare i bisonti verso la meta della caccia, ossia il punto del salto (*linea in colore*), che nel II secolo d.C. si trovava una ventina di metri al di sopra della scarpata di detriti.

nell'interpretazione di ciò che segue. Innanzitutto, studi sul paleoclima hanno stabilito che il periodo compreso grosso modo fra il 5500 e il 3000 a.C. fu «altitermico», ossia vide condizioni generalmente più secche e un po' più calde di quelle tipiche tanto prima che dopo. Poiché le praterie sono semiaride anche oggi, si pensò per molto tempo che gli americani più antichi avrebbero dovuto evitare le pianure in quel periodo di maggiore aridità. A quanto pare non fu però così. A prescindere dai reperti trovati a Head-Smashed-In, resti culturali sono stati trovati nel frattempo in varie gole del Wyoming in cui si intrappolavano bisonti fra il 4700 e il 4000 a.C. Tali resti appartengono alla tradizione culturale della Mummy Cave. Altre prove del fatto che il periodo caratterizzato da temperature elevate non fu nelle praterie un periodo di inattività sono state trovate in siti del Crowsnest Pass nelle Montagne Rocciose, 90 chilometri a ovest di Head-Smashed-In. Qui una serie di uccisioni di bisonti, portate a effetto inseguendo le mandrie in terreni paludosi, sono state datate al periodo compreso fra il 5200 e il 2800 a.C.: un intervallo di tempo che si inserisce fra il tardo Paleoindiano, prima dell'inizio del periodo caldo e 200 anni prima del ritorno, nelle praterie, delle attuali condizioni climatiche.

La notazione geologica potrebbe sorprendere coloro che pensano alle Great Plains come a grandi estensioni piatte e uniformi. E così sono, se si prescinde dalla morfologia associata all'erosione fluviale, dagli occasionali rilievi collinari e dalle catene montuose delle praterie. La piattezza del paesaggio, nella misura in cui appunto è piatto, è il prodotto della glaciazione pleistocenica che ricoprì il paesaggio preglaciale di una spessa coltre di depositi morenici glaciali. Questi depositi sono virtualmente un insieme di materie prime potenziali per utensili in pietra: quarzite, selce e ciottoli. Inoltre, almeno nell'Alberta meridionale, il substrato roccioso del Cretaceo sottostante al detrito glaciale contiene quarzite e conglomerati di selce, modeste quantità di legno pietrificato e, in una formazione (l'argillite di Bearspaw), quantità di ciottoli di selce nera. Le cave di selce nell'area del Knife River, nel North Dakota, sono un'altra area ben nota di materia prima per utensili. Esse forniscono una selce del colore del miele che fu molto ricercata e commerciata su grandi distanze.

Le Montagne Rocciose, che formano il confine occidentale delle praterie, offrono tipi di materiali per utensili sia simili sia del tutto diversi. Sono comuni selci di alta qualità, ma anche letti siltosi metamorfosati e noduli di calcedonio. La materia prima per utensili più tipica della regione sono le ossidiane. La maggior parte di esse provenivano dall'area che ora costituisce il Parco nazionale di Yellowstone, e l'ossidiana di Yellowstone veniva commerciata sino a siti della cultura di Hopewell a est del Mississippi.

Tanto basti sull'ambiente. Per tornare alla popolazione della fase della Mummy Cave a Head-Smashed-In, essa usava come materia prima per le punte di armi da lancio la quarzite locale. Da ciottoli di quarzite si preparavano nuclei da cui venivano staccati pezzi, ulteriormente lavorati per dar loro la forma di punte per armi da lancio; le punte erano bifacciali (lavorate su entrambe le facce), con un incavo su ciascun lato. Le loro dimensioni fanno pensare che fossero montate sul

Il dirupo di arenaria a Head-Smashed-In si erge una decina di metri al di sopra della scarpata di detriti al margine sudorientale delle Porcupine Hills, una regione di terreno ondulato. Gli indiani delle Great Plains appartenenti alla cultura detta della Mummy Cave potrebbero essere stati i primi a uccidere i bisonti facendoli cadere sulla scarpata sottostante. Altri tre gruppi di cacciatori di bisonti, compresi i predecessori degli indiani piegan, che diedero al sito il nome, usarono dopo di loro la stessa tecnica. Lo scavo sul pendio è stato eseguito dall'autore e collaboratori nel 1965, quando il Glenbow-Alberta Institute e l'Università di Calgary iniziarono una ricerca nel sito.

L'area di sondaggio principale era una fossa di tre metri per sei, che si restrinse poi a un pozzo più stretto sino a una profondità di otto metri. I materiali estratti dal pozzo, che raggiunse infine una profondità di 11 metri, venivano portati in superficie per mezzo della gru improvvisata che si vede qui in azione e venivano setacciati alla ricerca di reperti. Le ossa di bisonte provenienti dal livello a otto metri di profondità, analizzate al radiocarbonio, rivelarono un'età di circa 3000 anni.

tipo di giavellotto che veniva lanciato con un propulsore, una sorta di asta (chiamata *atlatl* dagli aztechi), che serve a imprimere maggior velocità al giavellotto. Gli uomini della cultura della Mummy Cave staccavano dai blocchi di quarzite anche grandi schegge, dalla forma irregolare, ma evidentemente produssero ben pochi dei coltelli in uso per scuoiare gli animali così tipici delle culture successive; ne trovammo infatti solo uno. Essi non erano, inoltre, molto interessati a materie prime non comuni per la fabbricazione di utensili. Soltanto l'un per cento delle loro punte era fatto di una pietra importata, e precisamente di selce del Montana.

Sopra i quattro strati corrispondenti alla cultura della Mummy Cave c'era uno strato sterile di detriti trasportati dall'acqua e di sabbia accumulata dal vento, il cui spessore variava da un po' più di mezzo metro a quasi due metri. Un suolo erboso si era sviluppato al di sopra di questo strato, facendo pensare che il «salto dei bisonti» non fosse stato usato attivamente per qualche secolo. Al di sopra dello strato sterile c'era un deposito di ossa di altezza compresa fra uno e due metri. Con un'eccezione anomala, le date ottenute per questi strati con il metodo del radiocarbonio variavano fra il 1090 a.C., con un'approssimazione in più o in meno di 120 anni, al fondo e il 25 d.C., con un'approssimazione in più o in meno di 80 anni, per la parte superiore. Tutti e quattro gli strati rappresentavano una cultura di cacciatori di bisonti delle pianure settentrionali e delle Montagne Rocciose nota come Pelican Lake, da un sito nel Saskatchewan meridionale.

Fra le ossa di bisonte furono trovate punte di armi da lancio tipiche della cultura del Pelican Lake, con incavi agli angoli. Esse erano state ricavate da schegge in forma di lama staccata a percussione da nuclei conici o poliedrici di selce di buona qualità. Gli uomini della cultura del Pelican Lake, a differenza dei loro predecessori, sapevano apprezzare materie prime importate per la fabbricazione di utensili. I loro coltelli (grandi e bifacciali), raschiatoi e utensili più piccoli con schegge ritoccate erano fatti di selce e calcedonio delle Montagne Rocciose, di ossidiana di Yellowstone e della selce color miele proveniente dalle cave del Knife River nel North Dakota.

Alcune fra le punte di armi da lancio degli strati del Pelican Lake, anziché avere un incavo agli angoli, sono peduncolate. Entrambi i tipi di punta furono fabbricati con una materia prima importata: argillite metamorfosata proveniente dall'area del Kootenay Lake nella British Columbia, a ovest di Head-Smashed-In. Sia la provenienza della materia prima sia la forma diversa di queste punte fanno pensare che visitatori provenienti dal Kootenay Lake abbiano partecipato occasionalmente alle cacce al bisonte della popolazione del Pelican Lake.

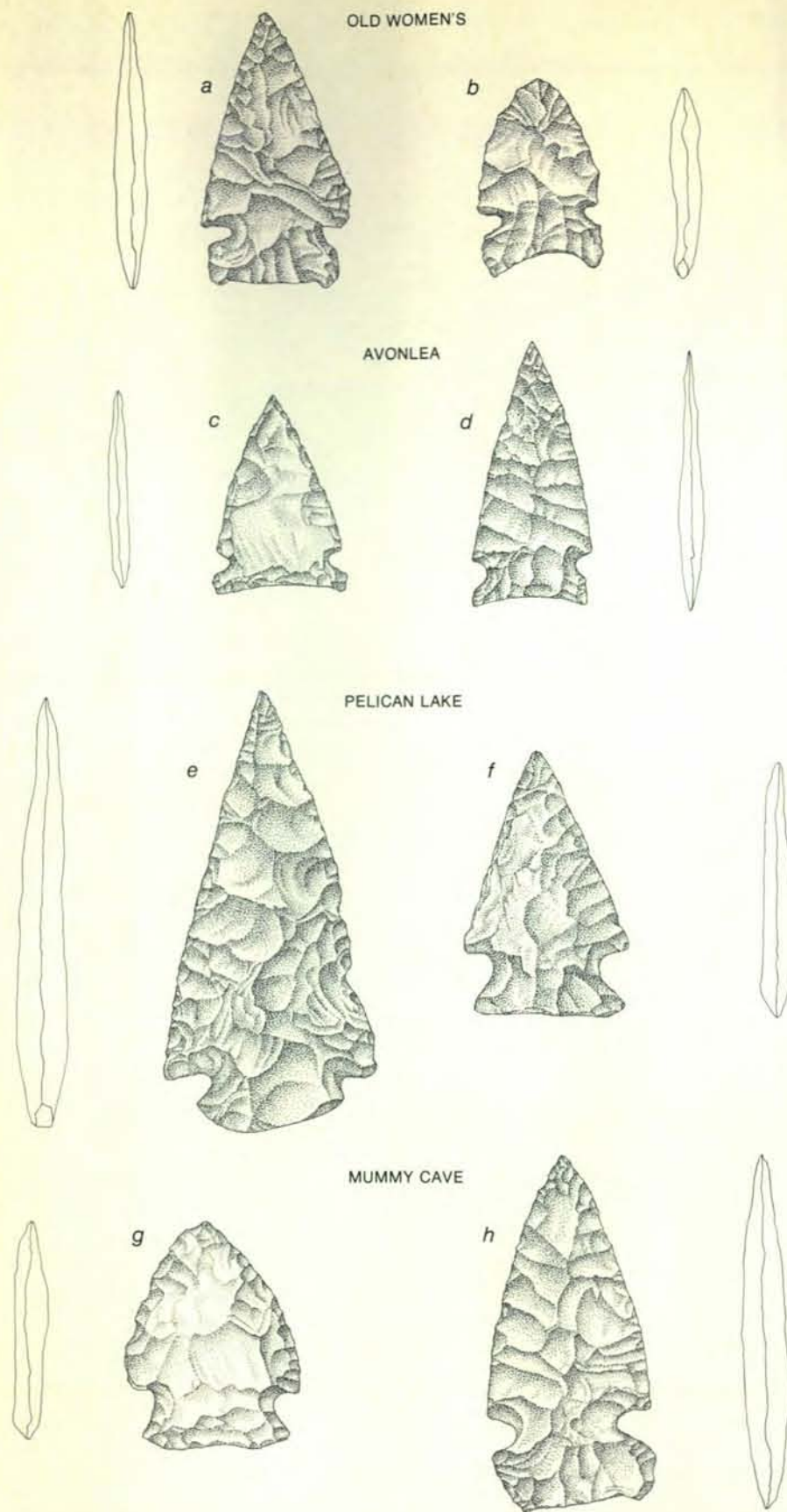
Un attributo che gli uomini del Pelican Lake avevano in comune con i loro predecessori fu l'uso dell'*atlatl* nella caccia. Fra le ossa di bisonte sono state trovate punte sia rotte sia integre di giavellotto. Le aste di giavellotto erano però piuttosto rare, probabilmente perché i giavellotti di solito dovevano essere recuperati e riportati all'accampamento per essere riarmati. Tale recupero era del resto giustificato: l'asta di un giavellotto, più grande di quella di una freccia e costruita probabilmente con cura molto maggiore, è l'elemento critico nel sistema balistico dell'*atlatl*, esattamente come l'arco lo è nel sistema dell'arco e freccia. Come vedremo, una volta che l'*atlatl* fu sostituito dall'arco, i cacciatori considerarono le frecce e le punte di freccia materiali sprecabili, a differenza dell'arco.

La seconda caduta di blocchi di arenaria ebbe luogo attorno al 100 d.C. Nella sezione meridionale della scarpata, grandi pezzi di arenaria si seppellirono nello strato superiore del Pelican Lake. A nord un'intera sezione del dirupo roccioso, di spessore compreso fra 15 e 21 metri, cadde percorrendo una distanza di 12 metri, ruotò e poi scivolò lungo il substrato roccioso inclinato, sospingendo verso valle una parte degli strati sovrastanti del Pelican Lake. Benché a quanto sembra la caccia al bisonte non abbia subito gravi interruzioni in conseguenza di questo evento catastrofico, pare che la metà settentrionale della scarpata di detriti sia stata poco usata in seguito. L'impatto iniziale del blocco franato aveva causato una profonda depressione; i cacciatori la sfruttarono come una sorta di recinto naturale, sospingendo i bisonti nel precipizio in modo che la maggior parte degli animali morti e feriti andassero a cadere nella depressione. Fra le ossa di bisonte accumulatesi qui trovammo per la prima volta punte di freccia.

Le piccole punte tipiche di freccia si annunciano a Head-Smashed-In nella fase di Avonlea. Questa fase prende il nome da un sito nel Saskatchewan, e gli uomini di Avonlea erano abili cacciatori di bisonti, ugualmente abili a sospingere gli animali in un precipizio o a guidarli all'interno di un'area chiusa o recintata. Questa cultura potrebbe essersi evoluta a partire dal complesso stesso del Pelican Lake. Almeno le prime punte di freccia di Avonlea a Head-Smashed-In sono punte con incavi agli angoli fatte con selce estratta dalle cave del Pelican Lake, nel Saskatchewan. L'evoluzione, se vi fu realmente, non ebbe luogo a Head-Smashed-In e nelle sue vicinanze. Fra tutte le punte che recuperammo dagli strati più bassi di Avonlea, quelle per giavellotto lanciato con l'*atlatl* costituivano meno dell'un per cento del totale.

Gli uomini della fase di Avonlea furono i primi, fra le popolazioni delle pianure, ad adottare l'arco e le frecce per la caccia. Essi vennero a conoscenza della nuova arma attraverso il contatto con popoli di pescatori nell'interno della British Columbia; questi a loro volta l'avevano appresa dagli indiani della Costa nordoccidentale, dove era in uso già da un millennio. A Head-Smashed-In gli uomini della fase di Avonlea furono così attivi nella caccia ai bisonti attuata facendoli precipi-

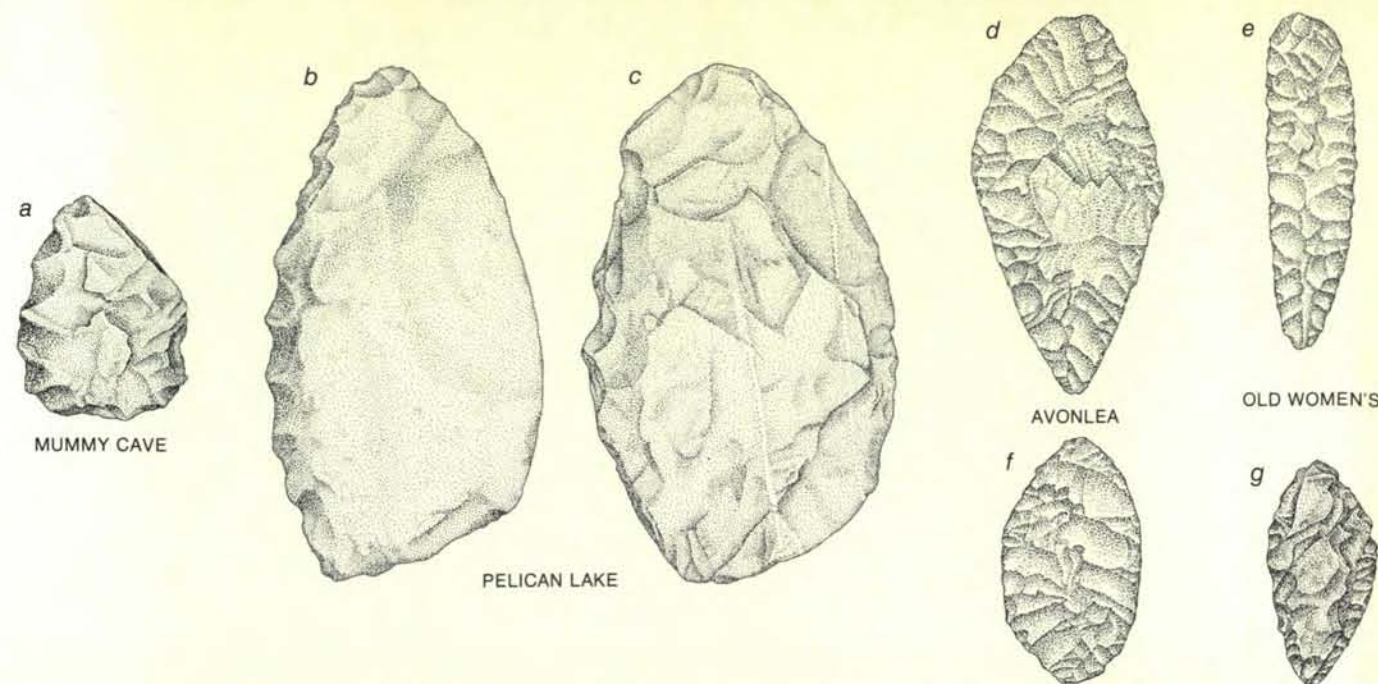
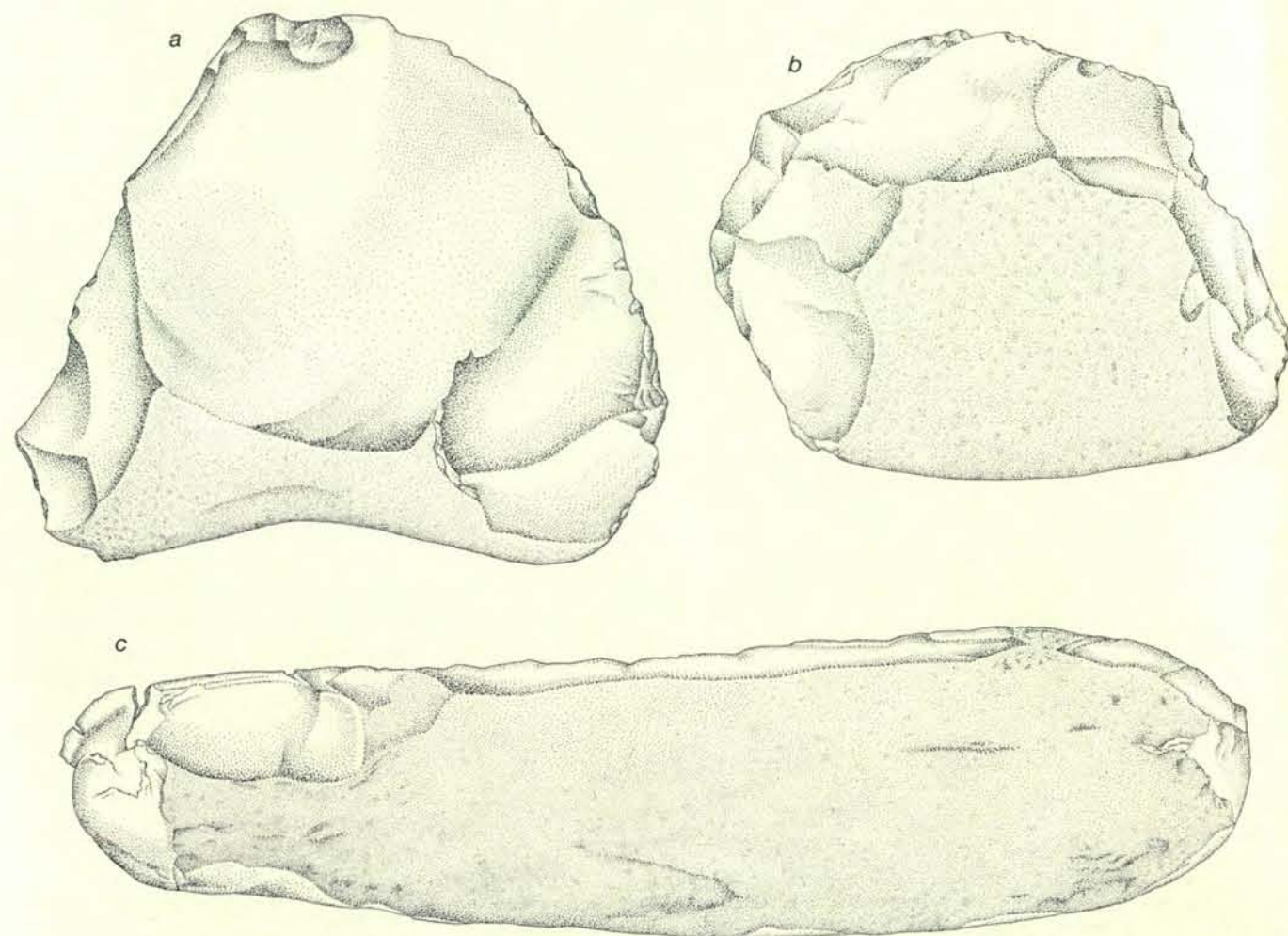
Si sta scavando con una cazzuola per riportare in luce altre ossa di bisonte nell'area orientale di Head-Smashed-In durante la stagione 1972. Sono già visibili un frammento di mascella e varie coste.

Questa buca nel sito dell'accampamento adiacente a Head-Smashed-In fu usata per estrarre per ebollizione il grasso dalle ossa di bisonte contenenti midollo. Una pelle di bisonte, con cui era rivestita la buca, veniva riempita d'acqua, la quale era poi portata a ebollizione immergendo in essa pietre previamente arroventate in un falò. Il grasso, recuperato e fatto raffreddare, era la base per la preparazione del *pemmican*: un cibo molto nutriente che poteva conservarsi per mesi.

Otto punte di armi da lancio appartenenti alle quattro fasi scoperte a Head-Smashed-In appaiono qui viste frontalmente e di taglio. Le quattro punte più in basso (*e-h*) sono punte di giavellotto da lanciare con un propulsore; la punta *e* è di ossidiana. Le quattro superiori (*a-d*) sono punte di frecce; il propulsore, l'*atlatl*, fu sostituito dall'arco nel sito studiato all'inizio della fase di Avonlea. Il popolo di Avonlea fu il più abile delle pianure nella lavorazione della pietra per distacco di schegge dopo la fine del periodo paleoindiano. La punta *d*, in ossidiana, è un buon esempio dell'abilità di questo popolo; rispetto alla punta *a*, pure di ossidiana, attesta una lavorazione molto più raffinata. Tutte le punte sono ingrandite del 50 per cento.

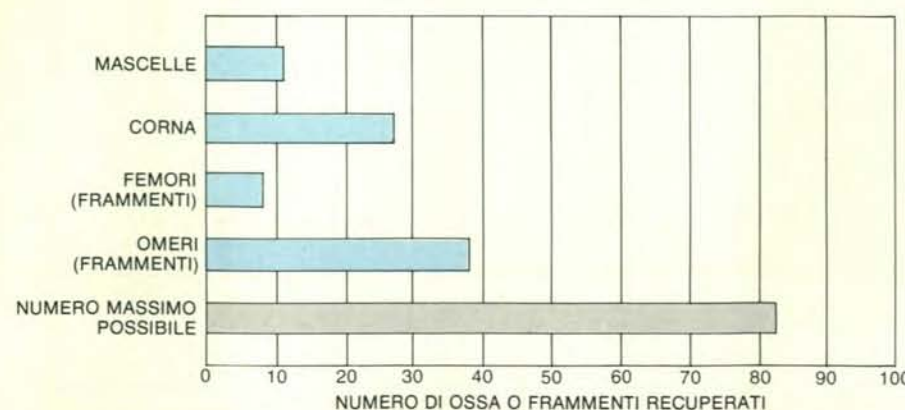
tare da un dirupo che il recinto naturale scavato dall'impatto dei blocchi di arenaria venne a contenere uno spesso strato di ossa. Nella parte inferiore del deposito di Avonlea, il cui spessore è di tre-quattro metri e che è datato attorno al 110 d.C., con un margine di approssimazione di più o meno 90 anni, le punte di freccia di Avonlea erano numerose. Molte di esse erano intatte; altre erano costituite solo dalla cuspide o dalla base della punta e dell'asticciola della freccia. A quanto pare gli arcieri di Avonlea non si curavano molto di recuperare le frecce dopo ogni battuta di caccia, forse perché molte aste di freccia si rompevano quando i bisonti precipitando colpivano il suolo.

Gli uomini di Avonlea importarono pochissima ossidiana di Yellowstone e niente selce delle cave del Montana come materia prima per le loro punte di freccia e per altri manufatti litici. Essi importarono invece selce di qualità eccezionalmente elevata dalla British Columbia e da altre fonti non ancora identificate. In ogni caso, lavorando selce importata, gli artigiani di Avonlea realizzarono i più raffinati manufatti in pietra che si fossero mai visti nelle Great Plains a partire dal tempo dei paleoindiani, circa 8000 anni prima. Le loro punte di freccia, fabbricate con piccole lame staccate per percussione da nuclei conici, sono estremamente sottili e furono lavorate delicatamente sino a dar loro la forma finale. Fra gli altri utensili tipici di Avonlea vi sono coltelli per scuoiare i bisonti, raschiatoi terminali, piccole schegge e cunei, questi ultimi usati forse per rompere le ossa degli arti dei bisonti allo scopo di estrarre il midollo. Dapprima lo sfruttamento delle pietre locali fu limitato alla fabbricazione di utensili robusti per tagliare, rompere e frantumare, ma col passare del tempo i fabbricanti di utensili di Avonlea vennero a dipendere sempre più dalle pietre delle Great Plains.

I livelli superiori di Avonlea, datati attorno all'850 d.C., presentano, rispetto a quelli inferiori, caratteristiche molto diverse. Qui le ossa di bisonte contenute nei consistenti depositi, di forma lenticolare, sono massicciamente carbonizzate e in alcuni casi completamente calcinate. Al di sopra di questi depositi c'è una serie finale di strati di ossa, alti due metri e mezzo, associati all'ultima fase culturale attestata a Head-Smashed-In: quella di Old Women's. Il nome è quello del sito vicino scavato nel 1958. Questa fase si estese dall'850 d.C. al presente storico (1800 d.C.), ovvero poco dopo il contatto di Peter Fidler con i piegan a Stimson Creek, verso nord-ovest. È probabile che la popolazione di Old Women's fosse costituita da piegan preistorici e che la sua cultura sia stata una fase evolutiva ulteriore della precedente fase di Avonlea. Gli strati di transizione fra queste due culture manifestano un mutamento graduale nello stile delle punte di freccia e di altri prodotti litici.

La fase di Old Women's è caratterizzata da piccole punte di freccia con incavi

Sette coltelli, rappresentativi delle varie fasi di Head-Smashed-In, sono qui illustrati a quattro quinti della loro grandezza naturale. Il numero dei coltelli rinvenuti nel sito è basso rispetto al numero delle punte di armi da lancio; negli strati della Mummy Cave ne è stato trovato solo uno (*a*). Tanto i coltelli di Avonlea quanto quelli di Old Women's, che erano dotati di un manico di legno, sono più raffinati di quelli più antichi.

Fra gli utensili litici pesanti, fatti principalmente di quarzite trovata nei depositi morenici locali, vi sono *chopper* ricavati da schegge (*a*) e da ciottoli (*b*). I *chopper* (utensili da percussione) conservavano di solito la forma naturale della pietra. Entrambi i tipi di utensile furono usati probabilmente più per rompere le ossa che per tagliare la carne. Le tracce di usura sull'utensile in forma di clava (*c*) fanno pensare che sia stato usato anch'esso per rompere ossa. La dimensione dei tre utensili qui raffigurati è circa due terzi delle dimensioni reali.

laterali profondi, da coltelli per scuoiare in forma di sigaro con immanicatura laterale di legno o d'osso, da raschiatoi terminali e da cunei, fatti tutti di ciottoli di selce locale. Il popolo di Old Women's, come i suoi predecessori, fece uso anche di grandi nuclei di quarzite e di schegge come *chopper,* utensili per frantumare ossa, e incudini e martelli per staccare schegge da piccoli ciottoli di selce allo scopo di ricavarne punte di freccia e coltelli. Nella fase antica la sua abilità nella lavorazione della pietra non era così raffinata come quella della popolazione di Avonlea, fatto che potrebbe essere attribuibile almeno in parte all'uso di selce locale di qualità non eccelsa. Furono prodotti anche manufatti di legno pietrificato. Verso la fine della fase di Old Women's, però, i produttori di utensili ridussero la loro dipendenza dalle materie prime delle Great Plains e importarono ossidiana da Yellowstone e porcellanite dal Montana meridionale. Un terzo tipo di materiale importato, spesso usato in fasi anteriori di Head-Smashed-In, la caratteristica selce color miele del Knife River, fu raramente utilizzato dai produttori di utensili di Old Women's.

Questa determinazione della composizione di una mandria di bisonti si fonda su un'analisi di un deposito di ossa a Head-Smashed-In (Avonlea Nord *2B*), dove le ossa rappresentavano 41 individui. Come indica il grafico superiore, erano in prevalenza animali giovani. L'età stimata alla morte, come si osserva nel grafico inferiore, colloca sette degli animali nella categoria degli individui pienamente adulti e solo tre nella categoria di quelli vecchi (di oltre 9,5 anni). Altri sette avevano meno di 2,5 anni e ventiquattro avevano un'età compresa fra 2,5 e 5,5 anni. Poiché le femmine di bisonte partoriscono di solito in maggio, questi animali devono essere stati uccisi in un periodo compreso tra quattro e cinque mesi dopo: dalla metà di settembre alla fine di ottobre.

Le ossa conservate nello stesso strato di Avonlea, Nord *2B*, dimostrano che, come fu d'uso anche in seguito, i cacciatori di Avonlea trasportavano molte parti dei bisonti uccisi lontano dal luogo dell'uccisione per procedere a un ulteriore sezionamento delle carni. La presenza di 41 individui fu stabilita mediante l'analisi dei denti. Se le carcasse fossero rimaste intatte, si sarebbero dovute trovare 82 corna, 82 fra mandibole e mascelle, e le ossa di 82 arti anteriori e di 82 posteriori (*istogramma in basso*). Si trovarono invece meno di metà di quel numero di corna e solo 11 fra mandibole e mascelle. I frammenti di femore erano solo otto. C'erano però 38 frammenti di omero, il che fa pensare che questi ultimi non venissero altrettanto rotti per l'estrazione del midollo.

Gli strati di ossa di bisonte della fase di Old Women's variarono considerevolmente per composizione. Gli strati, dello spessore di un metro, dei depositi inferiori comprendono mucchi di ossa carbonizzate frammiste a scheletri di bisonti non bruciati e ben conservati. Al di sopra di questo deposito inferiore c'era uno strato di ossa carbonizzate e non carbonizzate, di corna di bisonte e di pelo e pelle decomposti, mischiati al contenuto delle viscere di animali, comprese feci. Gli strati finali, dello spessore di 1,2 metri, contenevano ossa ben conservate, prive di tracce di fuoco, in una matrice friabile di colore rosso-arancione composta da pelle e pelo decomposti. Abbiamo trovato merci di produzione europea - punte di freccia in metallo - nello strato più alto, dalla superficie del quale abbiamo recuperato una palla di moschetto. Sulla base delle merci rinvenute qui e nel sito di accampamento vicino, assegnammo alla parte superiore del deposito la data del 1800 d.C. circa.

Il sito dell'accampamento, un ampio ripiano glaciale piatto 18 metri al di sotto della scarpata di detriti, ha un'area di sei ettari e mezzo. La maggior parte della sua superficie attuale è ricoperta da un sottile strato di rifiuti: ossa spezzate, focolari di roccia spaccata per cucinarvi cibi e manufatti di vario genere. L'area preferita per l'accampamento era evidentemente l'estremità meridionale del ripiano, sopravvento rispetto all'area in cui venivano uccisi i bisonti. Qui scoprimmo un mucchio di ossa, rocce spaccate e manufatti associati al sezionamento finale di varie parti dei bisonti uccisi. I principali manufatti compresi nel mucchio di rifiuti erano numerosi pesanti utensili in pietra.

Se i resoconti storici delle cacce al bisonte da parte dei piegan possono essere utilizzati come guida, i vari utensili in pietra devono essere serviti a sezionare gli animali in quattro quarti, a tagliare la carne in strisce per farle essiccare e a rompere le ossa degli arti per estrarne il midollo. La maggior parte dei nuclei di pietra e delle schegge di scarto trovati nel sito dell'accampamento erano stati ricavati da grandi pietre di quarzite a grana grossa; furono trovate anche piccole schegge di scarto di selce a grana fine. Venne alla luce anche una piccola quantità di cocci di

A Head-Shashed-In i manufatti litici di vario tipo variarono di numero da una fase all'altra. Negli strati della Mummy Cave, per esempio, dei 1129 manufatti trovati il 97 per cento erano schegge di quarzite a grana grossa ricavate da ciottoli di quarzite locale. Le punte di arma da lancio, 17 in tutto, dovevano essere fissate ai giavellotti che venivano lanciati con l'*atlatl*. Negli strati della Mummy Cave si trovò un solo coltello. Anche dei 1149 manufatti degli strati del Pelican Lake più della metà erano schegge di quarzite locale. Le punte di giavellotto da lanciare con l'*atlatl*, alcune delle quali di ossidiana importata, rappresentavano un altro 23 per cento del totale. Furono trovati anche coltelli, alcuni dei quali fabbricati con pietre importate, mentre attrezzi in pietra pesanti come *chopper* e pietre da percussione rappresentavano il 5,3 per cento del totale. Negli strati di Avonlea le punte di freccia erano le più abbondanti fra i 1735 manufatti trovati, rappresentando il 61 per cento del totale. I coltelli erano molto più numerosi di quanto non fossero nella fase del Pelican Lake, ammontando a 60 pezzi. Gli artigiani della pietra di Avonlea importavano selce dalla British Columbia e ossidiana dall'area di Yellowstone. La popolazione della fase finale, quella dell'insediamento di Old Women's, lasciò meno punte di freccia della popolazione precedente di Avonlea, ma un maggior numero di coltelli. Fino ai tempi storici essa lavorò principalmente la pietra locale, usando selce per punte di freccia e coltelli e quarzite per gli utensili pesanti.

ceramica, assieme a merci come perline per collane, braccialetti di rame, vetro di bottiglia e punte di freccia in ferro.

Fra le ossa di bisonte rinvenute nell'accampamento c'erano frammenti di mandibola, rimasti nel sito dopo l'asportazione della lingua dell'animale (considerata una leccornia dai cacciatori di bisonti sin dai tempi paleoindiani), ossa craniche basali e frontali, rimaste dopo l'estrazione del cervello, coste, ossa di gobba di bisonte e molti frammenti di ossa di arti. Scavammo anche un certo numero di quelle che, a giudicare dal loro contenuto e da resoconti storici, dovevano essere buche per l'ebollizione. Tradizionalmente tali buche venivano scavate sino a far loro assumere un diametro di due metri e una profondità di un metro. Una volta rivestita la buca con pelle di bisonte, la si riempiva d'acqua e di ossa con midollo rotte. Pietre fatte arroventare in un falò vicino venivano poi gettate nell'acqua fino a portarla all'ebollizione. Infine il grasso di midollo che si accumulava al di sopra dell'acqua veniva raccolto, versato in recipienti e usato in seguito come base per il *pemmican*: un miscuglio di grasso di bisonte, di bacche e di carne essiccata ridotto in polvere.

Benché sia probabile che l'area dell'accampamento sia stata usata da gruppi di cacciatori diversi dai piegan preistorici e moderni di Head-Shashed-In, trovammo poche o nessuna traccia archeologica delle due genti che avevano praticato la caccia al bisonte nel sito facendo precipitare gli animali dal dirupo. Head-Smashed-In è però solo il secondo sito in cui si praticava questo tipo di caccia che sia stato scavato nell'Alberta (il primo è quello di Old Women's). Almeno 20 altri siti, con i loro accampamenti di caccia e macellazione associati, rimangono ancora da scavare. Il nostro sito fornisce reperti che ne documentano l'uso da parte di cacciatori delle Great Plains per un periodo di almeno 5600 anni, un intervallo di tempo davvero considerevole. Solo un altro sito in cui si praticò la caccia a mandrie di bisonti spingendoli giù da un dirupo è sostanzialmente più antico: si tratta del Bonfire Shelter, in prossimità della confluenza del Pecos River con il Rio Grande nel Texas. Il Bonfire Shelter ha circa 10 000 anni, ma la storia del suo uso è molto più breve di quella dell'uso di Head-Smashed-In. Con l'eccezione del Bonfire Shelter, i *jump site* sono tipicamente distribuiti lungo il margine occidentale delle pianure. Non si sa quanto siano antichi altri *jump site* dell'Alberta.

In ogni caso le scoperte a Head-Smashed-In indicano che gruppi numerosi di cacciatori che cacciavano a piedi avevano successo nel procurarsi carne e grasso, pelli da usare nella confezione di indumenti e di tende, vesciche e tratti di intestino da usare come recipienti, tendini per cucire, ossa e corno per fabbricare utensili e attrezzi, e letame e sego da bruciare per ottenerne calore e luce. I loro metodi di caccia richiedevano inoltre meccanismi sociali più raffinati di quelli che molti studiosi sono stati disposti a riconoscere persino agli indiani delle praterie di epoca storica. Nonostante la loro cultura materiale ingannevolmente semplice, questi antichissimi uomini delle pianure erano infatti al tempo stesso eccellenti organizzatori e audaci coordinatori quando ancora molti fra i progenitori dei bianchi erano semplici cacciatori e raccoglitori nelle fredde foreste dell'Europa del Mesolitico.

L'estinzione delle ammoniti

I mutamenti subiti dalle conchiglie di questi animali marini, alla fine della loro storia, fanno pensare a una battaglia senza scampo che essi combatterono contro predatori più mobili e in grado di schiacciarli

di Peter Ward

Sessantacinque milioni di anni or sono, alla fine del Cretaceo, un periodo del Mesozoico, molte specie di piante e di animali, compresi numerosi gruppi di invertebrati marini, si estinsero improvvisamente. Fra gli invertebrati marini che allora scomparvero, le ammoniti costituiscono uno dei gruppi più importanti. Al pari dei rappresentanti del genere *Nautilus*, i loro parenti più vicini, oggi viventi, esse erano cefalopodi con conchiglia concamerata, ossia molluschi viventi nella cavità più esterna di una conchiglia che essi stessi, col procedere dello sviluppo, suddividevano in camere formando dietro di sé, con le proprie secrezioni, delle pareti divisorie o setti. Attualmente, molti ricercatori ritengono che le estinzioni verificatesi alla fine del Cretaceo siano state causate dall'effetto diretto o indiretto della collisione di un grosso corpo meteoritico con la superficie terrestre, un'eventualità proposta per la prima volta da Luis W. Alvarez, da Walter Alvarez e dai loro collaboratori dell'Università della California a Berkeley, per spiegare i livelli eccezionalmente elevati di metalli del gruppo del platino rinvenuti negli strati sedimentari risalenti a quel periodo (si veda l'articolo *L'estinzione in massa dei dinosauri* di Dale A. Russell in «Le Scienze» n. 163, marzo 1982). Per contro, le testimonianze fossili fanno pensare che la scomparsa delle ammoniti non sia stata la conseguenza di questa catastrofe, ma di variazioni rilevanti all'interno dell'ecosistema marino, nel tardo Cretaceo.

«Ammoniti» è la denominazione che distingue i rappresentanti della sottoclasse ammonoidei o ammonoidi del Giurassico e del Cretaceo. Le conchiglie delle ammoniti si differenziano da quelle tipiche degli ammonoidi per la maggior parte della storia di questo gruppo, prolungatasi per ben 300 milioni di anni in modo tale da far supporre che questi animali dai movimenti lenti e che forse frequentavano i fondali, dovessero affrontare pressioni selettive molto più forti del normale. La tipica conchiglia degli ammonoidi era adeguatamente strutturata per sostenere la pressione dell'acqua marina alla profondità a cui essi vivevano, ma non lo era in modo sufficiente per sfuggire o scoraggiare del tutto i predatori. La maggior parte delle conchiglie degli ammonoidi, come quelle del *Nautilus* attuale, avevano conformazione planispirale (cioè erano avvolte su un singolo piano) con pareti relativamente sottili e con forma scarsamente idrodinamica. Nel corso del tardo Mesozoico apparvero predatori marini, nuotatori veloci e in grado di frantumare le conchiglie, fra cui gli attuali pesci ossei (osteitti) e molte specie di rettili marini a respirazione aerea. La loro presenza può spiegare la prevalenza di tre tipi di conchiglie fra le ammoniti del Cretaceo: conchiglie planispirali idrodinamiche, conchiglie abbondantemente ornate o conchiglie planispirali corazzate e, infine, una grande varietà di forme eterogenee, definite collettivamente «eteromorfe».

Prendendo unicamente in considerazione le conchiglie, si possono fare molte ipotesi sull'agilità, sulla velocità e sui modi di vita delle nuove specie di ammoniti, ma per comprendere esaurientemente le strategie difensive di cui questi tipi di conchiglie sono taciti testimoni è necessario conoscere meglio la biologia degli animali che le abitavano. Nella maggioranza dei casi si sono conservate solo le conchiglie e altre parti solide delle ammoniti e solo raramente sono stati rinvenuti fossili contenenti tracce di materiale organico. Pertanto, il *Nautilus*, un vero fossile vivente, si è dimostrato una fonte d'informazioni insostituibile per chiarire modificazioni altrimenti inspiegabili o chiaramente non adattative delle conchiglie delle ammoniti.

Non è stato ancora spiegato come si è concluso il ciclo delle ammoniti, in parte anche perché i sedimenti che comprendono la fase transitoria fra Cretaceo e Terziario e che le contengono sono stati scoperti solo in pochi siti. Nel corso del tardo Cretaceo sia il numero delle ammoniti nel loro complesso sia quello delle loro specie avevano subito un brusco calo. Non è chiaro che cosa sia accaduto alle rimanenti e apparentemente più duttili specie di ammoniti, ma gli studi sui fossili contenuti nelle sezioni stratigrafiche di Zumaya, in Spagna, fanno pensare che esse si fossero estinte molto prima dell'ipotetico impatto del meteorite. Anche supponendo che questo impatto abbia accelerato, più che causato, l'estinzione definitiva delle ammoniti, l'ipotesi si presta a varie obiezioni. Alcune specie di ammoniti del Cretaceo assomigliavano molto ai nautiloidei o nautiloidi, gli antenati dei nautili attuali. Perché esse avrebbero dovuto soccombere mentre i nautili sono sopravvissuti?

I nautiloidi, di cui alcune specie sono state rinvenute in strati che risalgono al Cambriano, un periodo del Paleozoico, sono i cefalopodi più arcaici. Durante il Siluriano, un altro periodo del Paleozoico, si è sviluppato da questa sottoclasse un gruppo molto limitato (in senso numerico): i bactritidi. Questi potrebbero essere stati gli antenati comuni di tutti i cefalopodi più recenti, compresi gli attuali calamari, seppie e polpi. Probabilmente gli ammonoidi si sono diramati dai bactritidi, circa 395 milioni di anni or sono, agli inizi del Devoniano, un altro periodo ancora del Paleozoico.

La lunga storia evolutiva degli ammonoidi è caratterizzata da radiazioni improvvise e dalla comparsa subitanea di molte nuove specie, seguite rapidamente da brusche estinzioni. Durante il Paleozoico il numero delle specie di ammonoidi aumentò velocemente, eguagliando in breve tempo il numero delle specie di nautiloidi. Nel Triassico, un periodo del Mesozoico, e successivamente nel Giurassico e nel Cretaceo, gli ammonoidi ebbero una radiazione veramente esplosiva, eclissando i nautiloidi per numero di specie. Almeno tre quarti delle circa 10 000 specie fossili conosciute di cefalopodi con conchiglia concamerata sono ammonoidi. Questi sono talmente diversificati da servire spesso da utile riferimento per datare gli strati sedimentari.

Questa sezione trasversale di ammonite fossile mette in evidenza le camere interne della conchiglia che l'animale ha costruito, formando dietro di sé delle pareti durante le successive fasi della crescita. (Nel campione manca la cavità che contiene il corpo.) I cristalli di minerale siliceo (calcedonio e quarzo) depositatisi sulle pareti nel corso della fossilizzazione hanno prodotto l'effetto di stratificazione. L'esemplare è un membro del genere *Placenticeras*, uno dei generi di ammoniti che si sono estinti per ultimi. La forma della conchiglia è simile a quella del nautilo attuale, l'unico cefalopodo a conchiglia concamerata vivente. Uno degli interrogativi posti dall'estinzione delle ammoniti è perché generi come *Placenticeras* siano scomparsi, mentre gli antenati del nautilo sono sopravvissuti. Il termine ammonite si riferisce solo agli ultimi rappresentanti del gruppo degli ammonoidi. Le circa 7500 specie note di questi cefalopodi presentano conchiglie molto differenziate per forma e dimensioni. Il nome ammonoidi deriva dall'antica divinità egizia Ammon, spesso raffigurata con il capo di un ariete, le cui corna avvolte e a spirale e costolute assomigliano alle conchiglie degli ammonoidi. La maggior parte delle ammoniti aveva dimensioni minori di questo esemplare, che ha un diametro di 24 centimetri, ma sono stati anche trovati fossili con un diametro di quasi due metri.

La differenza più marcata fra le conchiglie degli ammonoidi e quelle dei nautiloidi consiste forse nei setti (pareti delle camere della conchiglia) e nelle relative suture (linee di contatto fra i setti e la parete interna della conchiglia), i due aspetti della conchiglia che ne determinano prevalentemente la robustezza. Nei nautili attuali, le camere vengono svuotate del liquido al momento della formazione di un nuovo setto, un procedimento che permette all'animale di ottenere un galleggiamento indifferente (cioè densità, o peso in assoluto per unità di volume quasi pari a quella dell'acqua marina circostante) e di poter così rimanere a qualsiasi profondità senza sforzo alcuno. Il peso della conchiglia e dei tessuti corporei supera quello dell'acqua marina, cosicché le camere sono soggette a pressioni molto più basse (meno di un'atmosfera) della pressione dell'acqua marina (circa 40 atmosfere) alla profondità dove abitualmente vive il nautilo (400 metri). Le conchiglie dei nautiloidi e degli ammonoidi erano anche probabilmente riempite, almeno in parte, di gas.

I setti delle conchiglie funzionano da sostegni, fornendo la forza sufficiente per controbilanciare la pressione dell'acqua alle grandi profondità. Nella maggior parte delle conchiglie di nautiloidi, essi sono superfici lisce leggermente incavate verso la camera più interna. Le suture all'intersezione fra setti e parete della conchiglia sono ispessite, ma le linee che formano sono curve semplici. Nella maggior parte degli ammonoidi la periferia dei setti è scanalata o corrugata e le linee di sutura sono increspate o dentellate e sorprendentemente complesse in confronto a quelle dei nautiloidi. Ammonoidi e nautiloidi si trovano comunemente negli stessi strati sedimentari ed è quindi probabile che siano vissuti alle stesse profondità negli oceani. Le pareti e i setti delle conchiglie degli ammonoidi sono però molto più sottili di quelle dei nautiloidi dello stesso diametro. È quindi evidente che la loro complessità strutturale le rendeva altrettanto forti delle conchiglie dei nautiloidi, pur essendo più leggere.

Forse gli ammonoidi hanno finito per produrre tipi di conchiglie che i nautiloidi non sono mai riusciti a sviluppare: per esempio conchiglie a sezione trasversale molto compressa o depressa (cioè conchiglie con una bocca esagerata in altezza o in larghezza). Probabilmente non è un caso che la varietà delle forme di conchiglie e il numero delle specie di ammonoidi siano aumentati di pari passo con l'evoluzione di suture settali più complesse. La rapidità con cui la forma e l'ornamentazione della conchiglia cambiavano sta però a indicare che entrambi questi caratteri erano assai plastici e possono avere avuto un limitato significato adattivo. Per contro le linee di sutura rimanevano del tutto stabili, tanto da poter essere usate per determinare linee evolutive.

Anche il piano costruttivo di una conchiglia tipica degli ammonoidi fa ritenere che, per gran parte del corso della storia di questo gruppo, forma e ornamentazione non fossero state sviluppate unicamente per permettere all'animale di sfuggire o di scoraggiare i predatori. Tutti gli ammonoidi del Paleozoico e quasi tutti quelli del primo e medio Mesozoico presentavano conchiglie planispirali, ma molte di queste non erano sufficientemente idrodinamiche e quindi non erano adatte al nuoto veloce. Esse erano in gran parte lisce o con ornamentazioni molto fini e offrivano pertanto una scarsa protezione di fronte ai predatori.

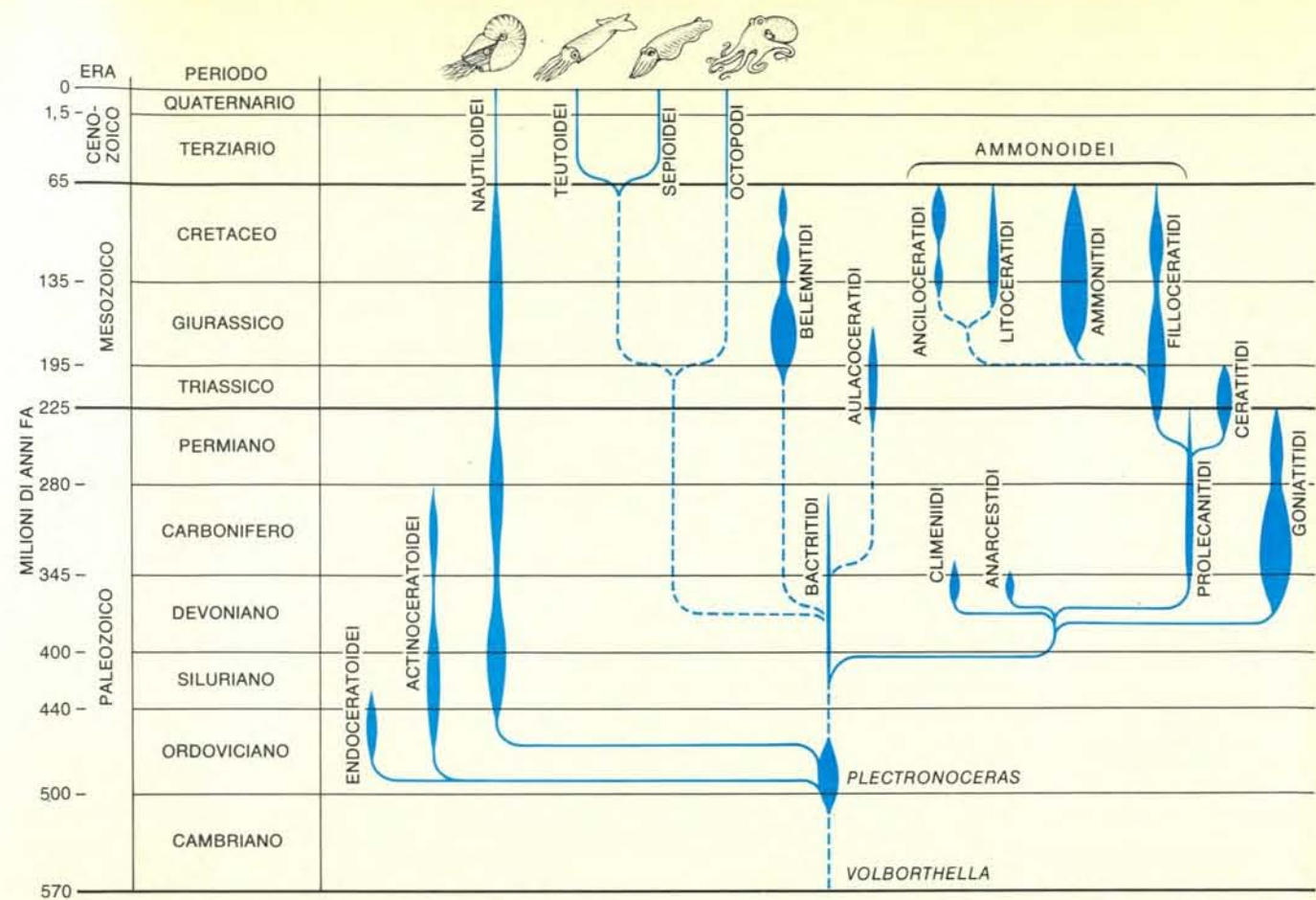
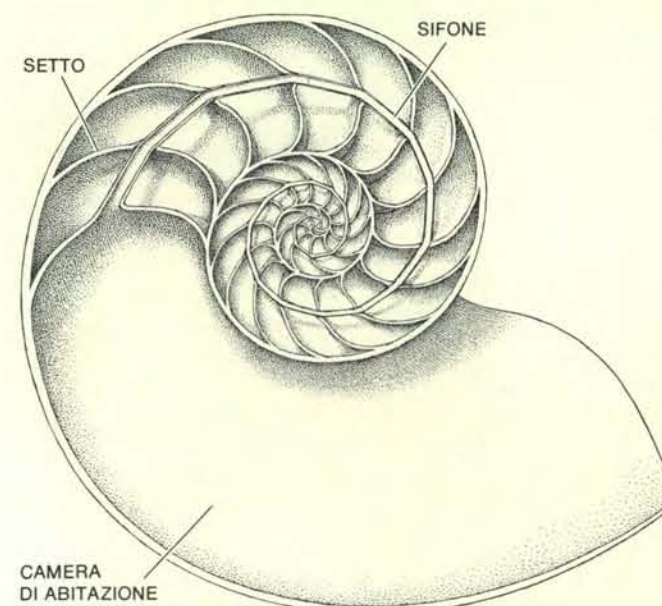
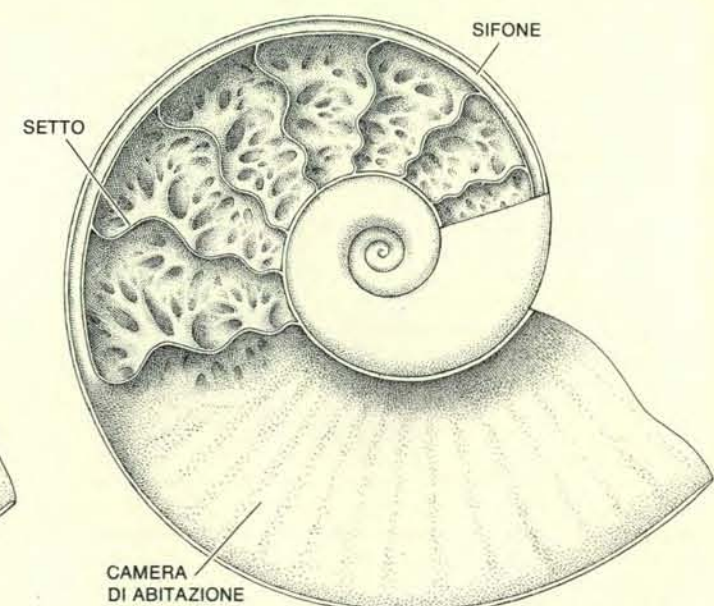
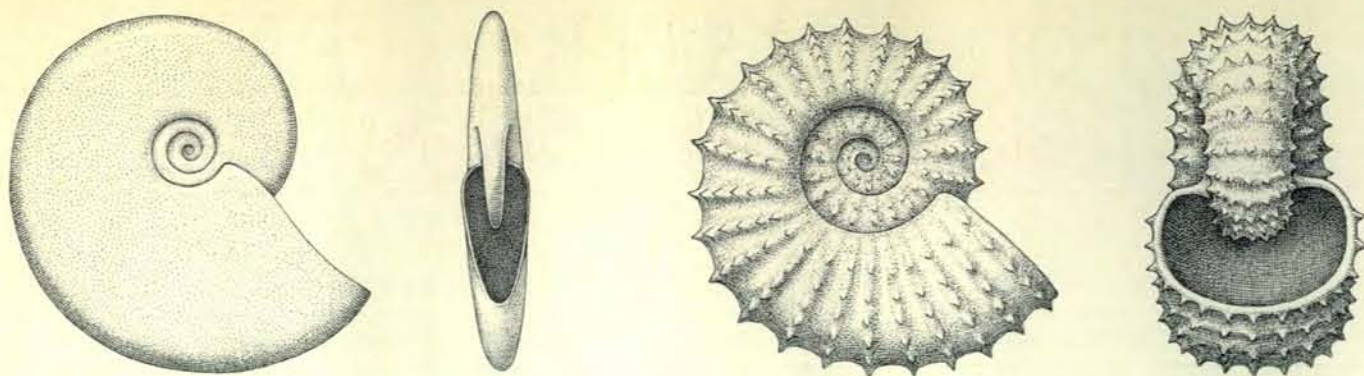
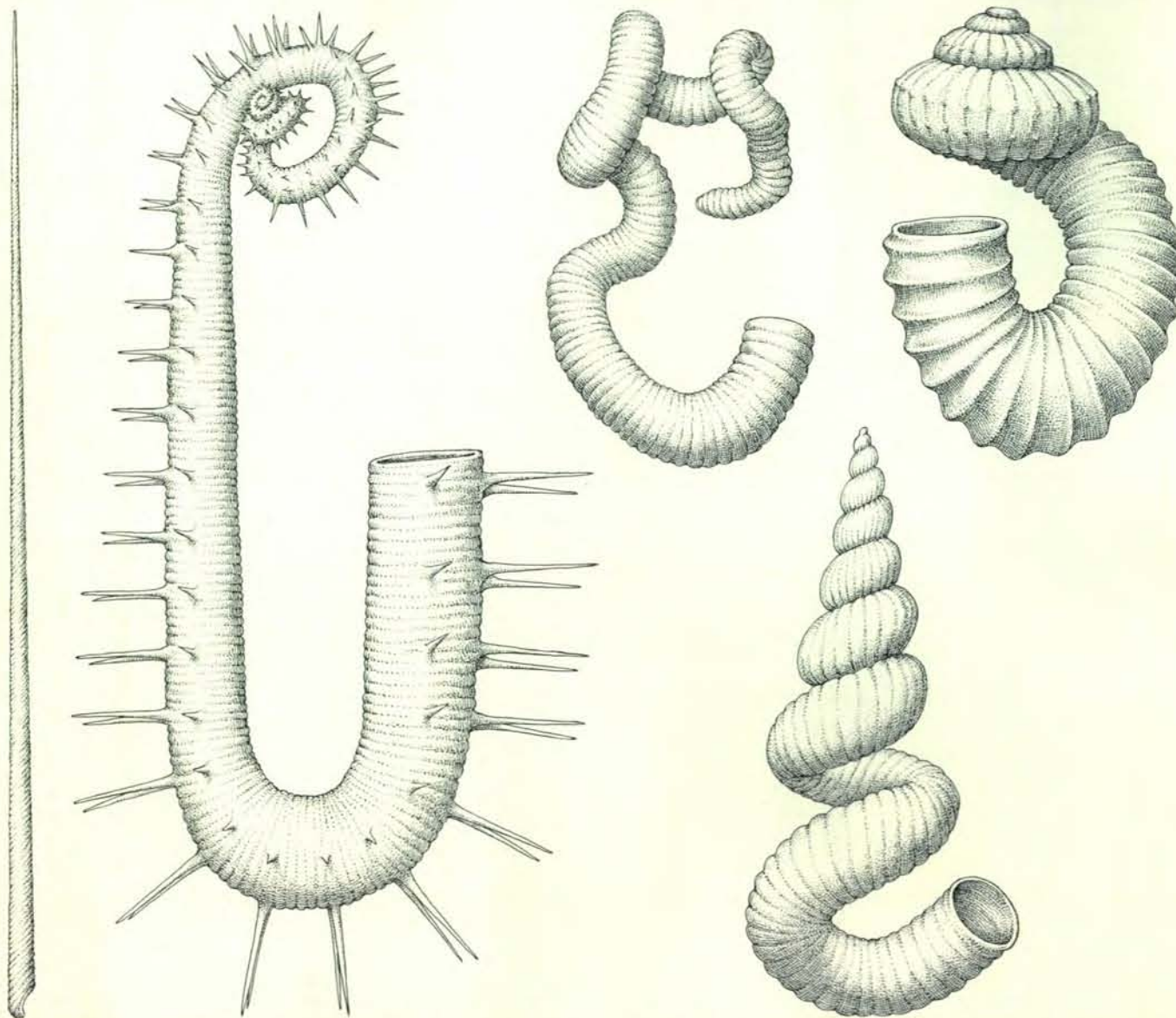
La situazione si trasformò nettamente durante il tardo Mesozoico quando, secondo Geerat J. Vermeij dell'Università del Maryland a College Park, si verificò la cosiddetta «rivoluzione marina del Mesozoico». Particolarmente durante il Giurassico e il Cretaceo, le radiazioni di animali in grado di frantumare conchiglie, come gli attuali granchi forniti di chele, le aragoste, le razze e le mante, unitamente alla comparsa di gasteropodi carnivori, di pesci osteitti e di molte specie di grossi rettili marini a respirazione aerea aumentarono sensibilmente il tasso e l'efficienza della predazione. Probabilmente sempre più spesso gli ammonoidi, essi stessi predatori di minuti animali marini, di piccoli crostacei e di ammonoidi di minori dimensioni, divennero prede o furono costretti a cedere il proprio bottino a predatori più mobili. Ciascuna delle tre tendenze tipiche nelle conchiglie delle ammoniti del Cretaceo è un chiaro adattamento sollecitato dalla pressione selettiva imposta dai predatori marini più efficienti.

La prima tendenza aveva prodotto forme di conchiglia che conferivano alle ammoniti maggiore agilità e la possibilità di nuotare più velocemente. Dal semplice studio della forma della conchiglia è possibile dedurre un sorprendente quantità di informazioni. Nel XIX secolo, il matematico inglese Henry Moseley sviluppò un modello geometrico dell'avvolgimento della conchiglia basato sulla spirale logaritmica, ossia su una spirale in cui la distanza dal punto di origine aumenta geometricamente a ogni rivoluzione. Nel corso degli anni sessanta, sulla base di quel lavoro, David M. Raup dell'Università di Chicago ha sviluppato delle equazioni matematiche che, variando quattro parametri della conchiglia, generavano una forma confrontabile con un modello esistente. Quando la conchiglia non è avvolta a elica, ma è planispirale, è sufficiente considerare solo tre parametri: la forma della curva chiusa che genera la conchiglia quando viene fatta ruotare attorno all'asse di avvolgimento, spesso espressa come rapporto fra larghezza e altezza della bocca della conchiglia (S); la proporzione in cui, a ogni rivoluzione, la curva generatrice si espande (W); la proporzione in cui, sempre a ogni rivoluzione, la curva generatrice si allontana dall'asse di avvolgimento (D). Per esempio, se la curva generatrice è costituita da una circonferenza, W è grande e D è piccolo, e il risultato sarà una conchiglia con avvolgimento molto serrato e volute con margini sovrapposti.

Raup ha continuato a rappresentare graficamente i campioni di ammonoidi sulla base dei valori S, W e D delle loro conchiglie e ha trovato che la struttura di molte di queste era scarsamente adattata per il nuoto. Gli adattamenti che hanno permesso agli ammonoidi (e ai nautiloidi) di nuotare liberamente invece di strisciare sui fondali sono stati due: lo sviluppo di un fragmocono, la parte della conchiglia svuotata dal liquido, che permette il galleggiamento indifferente, e la formazione di un iponoma, la struttura a forma di imbuto presente nella cavità corporea e attraverso la quale l'acqua può essere espulsa con forza.

Uno dei problemi che deve affrontare un ammonoide che nuoti riguarda il controllo della posizione. In un ammonoide il centro di galleggiamento, situato nel fragmocono, è separato dal baricentro, che si trova nella camera di abitazione. Pertanto questo animale ha una grande stabilità statica. D'altro canto, la forza di propulsione proveniente dall'iponoma è spostata rispetto al centro di galleggiamento e tende a far ruotare la conchiglia eccentricamente rispetto alla posizione di equilibrio; la forza opposta, prodotta dallo spostamento della massa, tende a riequilibrare la conchiglia. Quanto più il centro di galleggiamento è distante dal baricentro, tanto maggiore è la forza riequilibratrice e, di conseguenza, la stabilità dinamica della conchiglia. Quest'ultimo aspetto costituirebbe un impedimento all'agilità dell'animale, limitandone anche la velocità natatoria.

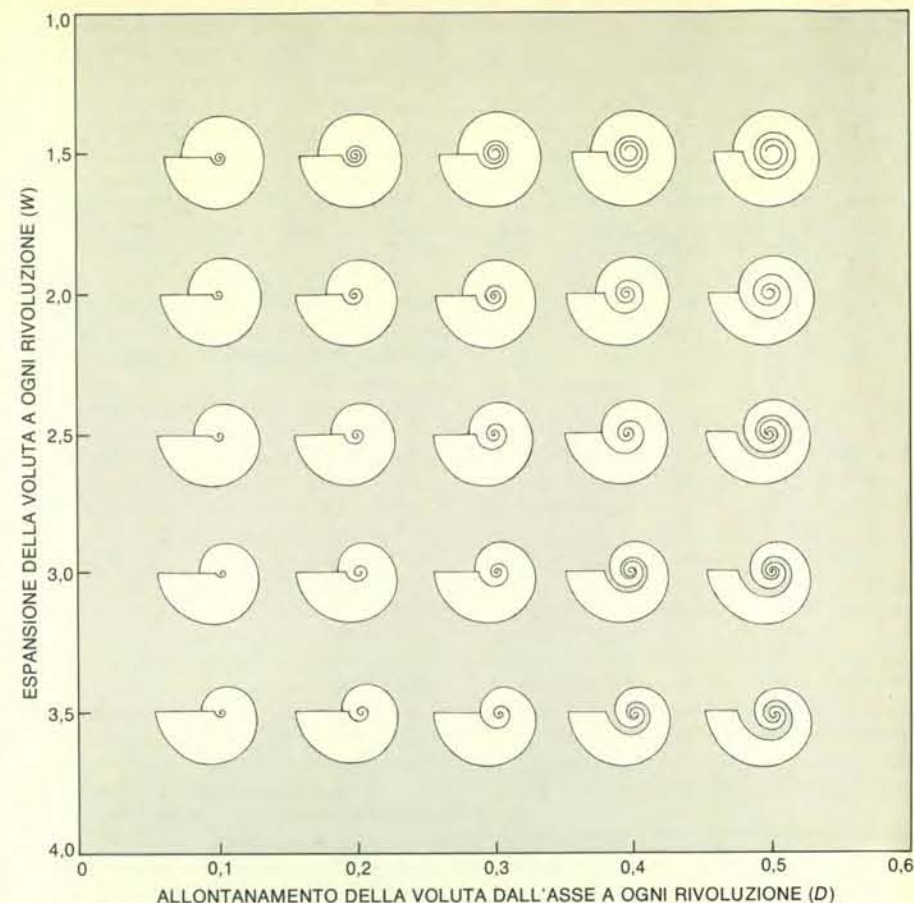
Un buon esempio di conchiglia in cui il centro di galleggiamento e il baricentro sono ampiamente separati è offerto dal nautilo. Raup ha trovato però che molte conchiglie di ammonoidi sono evolute, cioè formano una spirale larga (valore alto di W e basso di D) a differenza della conchiglia del nautilo, che è involuta, cioè forma una spirale stretta. In tale conchiglia il baricentro e il centro di galleggiamento tendono a essere ravvicinati. Di conseguenza, la spinta prodotta dall'espulsione dell'acqua dall'iponoma viene trasformata in un moto rotatorio attorno a un asse verticale o trasversale più che in un movimento in avanti.

John A. Chamberlain del Brooklyn College della City University di New York ha approfondito il lavoro di Raup sulle tipologie delle conchiglie di ammonoidi, con particolare riguardo alla struttura idromeccanica della conchiglia. Alla stabilità idrodinamica egli ha aggiunto il fattore supplementare della resistenza opposta dall'acqua quando una conchiglia è spinta in avanti o trascinata. Il parametro della conchiglia che ha l'effetto più marcato su tale resistenza è la forma della curva generatrice S. Se il rapporto fra larghezza e altezza della bocca della conchiglia è piccolo la curva è compressa e si ha una conchiglia a forma di disco; nel caso opposto, la curva è depressa e la conchiglia è globulare. Evidentemente, quanto più piccolo è il valore di S tanto minore sarà lo sforzo richiesto all'animale per nuotare a una determinata velocità.

Un piccolo gruppo di nautiloidi (antenati del nautilo attuale a conchiglia concamerata), denominati bactritidi, si trova probabilmente all'origine degli ammonoidi e delle belemniti, gruppi ambedue estinti, e di tutti i cefalopodi attuali, dei quali fanno parte le seppie, i calamari e i polpi. I piccoli bactritidi ortoconici (che cioè presentavano una conchiglia diritta) cominciarono a svilupparsi in ammonoidi quando le loro conchiglie iniziarono ad avvolgersi a elica dapprima in larghi giri e poi in giri sempre più stretti. Le belemniti possedevano una conchiglia concamerata interna rispetto al corpo molle. I cefalopodi attuali presentano o una conchiglia concamerata interna, un osso di seppia lamellare, un «gladio» più ridotto nel calamaro, oppure, nel caso del polpo, uno scheletro vestigiale consistente di piccoli supporti ossei o di una semplice asticciola. Queste parti solide interne sono omologhe alle conchiglie esterne concamerate dei nautiloidi e degli ammonoidi.

Le conchiglie degli ammonoidi differiscono per due aspetti da quelle dei nautiloidi. I setti (pareti delle camere) delle conchiglie dei nautiloidi hanno superfici lisce e incurvate. Le suture settali (linee di intersezione con la parete interna della conchiglia) sono leggermente curve. Invece i setti delle conchiglie degli ammonoidi sono scanalati alla periferia e le suture settali sono ripiegate a formare curve complesse sfrangiate. Appunto in ragione di questa complessità dei setti e delle loro suture, le pareti delle conchiglie delle ammonoidi erano relativamente sottili. L'altra diversità esistente fra ammonoidi e nautiloidi è la posizione del sifone, l'organo che provvede a svuotare le camere del liquido. Nei nautiloidi, il sifone generalmente passa al centro della voluta, mentre negli ammonoidi esso è spostato in direzione più periferica.

I tipi di conchiglie planispirali idrodinamiche e ornate sono illustrati da questi esemplari di ammoniti dei generi *Placenticeras* (*a sinistra*) e *Douvilleiceras* (*a destra*). La superficie esterna liscia e la sezione trasversale compressa della conchiglia di *Placenticeras* riducevano l'inerzia e l'avvolgimento molto stretto aumentava la stabilità idrodinamica; il nuoto dell'animale ne risultava più agile e veloce. La conchiglia di *Douvilleiceras* era molto meno adattata al nuoto a causa della grossolana superficie esterna, dell'ampia sezione trasversale, dell'avvolgimento in giri più larghi e della depressione dell'ombelico (la cavità che circonda l'asse di avvolgimento). Le spine e le coste ingrossate della conchiglia possono però aver contribuito a scoraggiare granchi, rettili marini come i mosasauri e altri predatori in grado di frantumare le conchiglie.

Le conchiglie eteromorfe delle ammoniti hanno talvolta forme così bizzarre da essere ritenute un tempo la conseguenza di un «esaurimento genetico», foriero dell'estinzione di questi cefalopodi. Anche oggi la varietà morfologica di queste conchiglie impedisce di compiere semplici generalizzazioni sul loro significato adattativo. Quelle qui illustrate appartengono (*da sinistra a destra*) ai generi *Sciponoceras, Anisoceras, Nipponites, Nostoceras* e (*in basso a destra*) *Didymoceras*. La maggior parte delle specie eteromorfe conduceva probabilmente una vita in prevalenza passiva galleggiando sull'acqua. Sembra, inoltre, che molte di quelle del primo Cretaceo dipendessero dall'ornamentazione per proteggersi contro i predatori; quelle del tardo Cretaceo, invece, e soprattutto quelle con un uncino a forma di *U* posto all'estremità della conchiglia, galleggiavano probabilmente a media profondità, discostandosi dalla piattaforma continentale, che aveva costituito gli habitat dei loro antenati bentonici e sfuggendo così forse anche ai predatori che avevano occupato quegli habitat nel corso del Mesozoico.

Chamberlain ha formulato equazioni che permettono di calcolare la velocità del nuoto degli ammonoidi in base a queste e ad altre caratteristiche delle conchiglie. La forma più adeguata per permettere all'animale di raggiungere la massima velocità consisterebbe in una conchiglia compressa con volute sovrapposte. Chamberlain ha trovato che conchiglie di ammonoidi, riprodotte graficamente in base ai parametri *S*, *W* e *D*, potevano essere raggruppate in forme che erano probabilmente le più adeguate al moto natatorio, ma che d'altra parte esisteva un gran numero di conchiglie corrispondenti a migliaia di specie di ammonoidi, che lo erano solo in minima parte o in modo insufficiente. Queste ricerche fanno pensare che, per gran parte della storia degli ammonoidi, la capacità di nuotare fosse stata scarsamente selezionata. Durante il tardo Cretaceo, invece, le forme molto idrodinamiche di conchiglie aumentarono a svantaggio di quelle che lo erano meno, sia regionalmente sia sul piano mondiale. Evidentemente, l'agilità e la velocità si erano dimostrate vantaggiose.

La seconda tendenza riguarda l'aumento dell'ornamentazione nelle specie di ammoniti con conchiglia non idrodinamica. Per la massima parte della loro storia gli ammonoidi hanno presentato conchiglie lisce o finemente ornate. Sebbene Chamberlain e Gerd E. G. Wastermann della McMaster University ritengano che qualche ornamentazione della conchiglia abbia effettivamente potuto ridurre la resistenza idrodinamica, come accade per le incavature delle palline da golf agli effetti della resistenza aerodinamica, molte ornamentazioni non sembrano rispondere a precise necessità. Inoltre, il tipo di ornamentazione è estremamente variabile all'interno di una determinata linea evolutiva. Sembrerebbe quindi che, a parte sostanziali affinamenti di schemi grossolani, il tipo di ornamentazione delle conchiglie di ammonoidi non abbia avuto particolare importanza e fosse assai plastico dal punto di vista evolutivo.

Nel Cretaceo, invece, quelle ammoniti planispirali che non presentavano conchiglie idrodinamiche svilupparono grosse e robuste coste, spine e tubercoli; nei molluschi attuali, sembra che questo tipo di ornamentazione abbia lo scopo di frustrare gli attacchi di predatori in grado di schiacciare le conchiglie. La tendenza allo sviluppo di conchiglie pesantemente corazzate fra le ammoniti può essere probabilmente interpretata allo stesso modo.

Questa particolare strategia difensiva che grado di efficacia aveva? Molti fossili di ammoniti del Cretaceo portano i segni degli attacchi dei predatori. Alcune conchiglie sono state aperte dai granchi e altri fossili mostrano tracce dei denti di mosasauri e di altri rettili marini. In molti casi, la conchiglia è stata spezzata nei pressi della bocca. Anche supponendo che l'animale non fosse stato ucciso, come avrebbe potuto sopravvivere a un simile attacco? Ancora una volta, gli studi sui nautili hanno fornito i necessari indizi.

Le capacità natatorie degli ammonoidi planispirali possono essere dedotte dalla forma delle loro conchiglie. David M. Raup dell'Università di Chicago ha sviluppato alcune equazioni che, variando quattro parametri della conchiglia, possono dare origine a qualsiasi forma di conchiglia basata su una spirale logaritmica. Queste figure tracciate dal calcolatore mostrano la variazione della forma della conchiglia in conformità a due parametri: in quale proporzione la sezione trasversale della conchiglia aumenta a ogni rivoluzione (*W*) e in quale proporzione la sezione trasversale si discosta dall'asse di avvolgimento a ogni rivoluzione (*D*). Le conchiglie con basso indice *W* e alto indice *D* (evolute) sono poco adattate al nuoto; quando i valori di questi indici si invertono, le conchiglie (involute) risultano relativamente bene adattate. Il baricentro delle conchiglie evolute, sito nella cavità della conchiglia dove è sistemato il corpo dell'animale, tende ad approssimarsi al centro di galleggiamento, all'interno del fragmocono, la parte non concamerata della conchiglia. Nelle conchiglie involute, i due centri sono più distanziati. La spinta prodotta dal getto direzionale d'acqua all'interno della camera d'abitazione che contiene il corpo dell'animale tende a far ruotare la conchiglia allontanandola dalla posizione di equilibrio (in cui il centro di galleggiamento è direttamente sovrastante il baricentro). Tanto più ampia è la separazione fra centro di galleggiamento e baricentro, tanto più velocemente la conchiglia è in grado di orientarsi. Pertanto, gli animali con conchiglia planispirale involuta tendono a essere nuotatori più agili e veloci.

Il fragmocono della conchiglia del nautilo serve principalmente a tenere a galla la pesante conchiglia calcarea più che la massa corporea. La densità della conchiglia è molto superiore alla densità media dei tessuti (2,67 grammi per centimetro cubo, in confronto agli 1,06 grammi dei muscoli e della ghiandola digerente e agli 1,03 grammi dei liquidi corporei). Eric Denton della Marine Biological Association of the United Kingdom di Plymouth ha calcolato che, per far galleggiare la massa della conchiglia, è necessario il concorso di circa l'80 per cento del volume del fragmocono. Di conseguenza, qualsiasi perdita di materiale della conchiglia produce una diminuzione relativamente grande della densità dell'animale. Nel caso di sufficiente asportazione di materiale, l'animale è costretto a risalire in superficie, esponendosi così ai predatori presenti in questa zona e rimanendo sballottato dalle onde e dalle correnti.

Molte conchiglie di nautili presentano fratture ricomposte, dal che si può dedurre che questi animali hanno una qualche capacità di sopravvivere a improvvise perdite di peso. La più ovvia strategia compensativa sarebbe quella di riempire di liquido le camere vuote. Fino a due anni or sono, però, si pensava che un nautilo in superficie o nei suoi pressi non avesse altra scelta che svuotare in maggiore o minore misura le camere, non quella di riempirle.

Denton e John B. Gilpin-Brown della Marine Biological Association hanno spiegato in che modo il nautilo svuota le camere chiuse. Studiando cefalopodi oggi viventi, a conchiglia concamerata come il nautilo, la seppia (che possiede un osso

concamerato interno) e *Spirula* (un cefalopodo affine alla seppia, che presenta una piccola conchiglia concamerata interna), essi si sono resi conto che le camere delle conchiglie o dell'osso di seppia sono originariamente riempite da un liquido salino denominato liquido camerale e non da gas, il che ha favorito una nuova interpretazione del sifone, l'organo che provvede all'evacuazione delle camere. Il meccanismo utilizzato dal sifone per quest'operazione è stato identificato da Denton, Gilpin-Brown, Lewis Greenwald e da me.

Nel nautilo, il sifone è costituito da un filamento di tessuto vivo racchiuso in un tubo calcareo che attraversa avvolto a spirale tutte le camere della conchiglia, compresa la prima che si forma con la crescita dell'animale. Le sostanze disciolte nel liquido camerale attraversano gli strati porosi esterni del sifone riversandosi nelle cellule del suo epitelio. Gli enzimi presenti sulla membrana interna pieghettata di queste cellule pompano i soluti nei fini reticoli delle pieghe; quando la concentrazione di ioni all'interno di queste pieghe è sufficientemente alta, il liquido proveniente dalle camere viene assorbito per osmosi passando successivamente ai vasi sanguigni situati al centro del sifone.

In quel periodo, non avendo mai potuto osservare un nautilo in superficie o nei suoi pressi mentre riempiva le proprie camere, avevamo ritenuto che la porzione di sifone che si estende fino alle prime camere che si svuotano del proprio liquido fosse il residuo vestigiale di un organo dei cefalopodi ancestrali a conchiglia concamerata, in grado sia di riempire sia di svuotare le camere. In seguito, Greenwald e io abbiamo dimostrato che anche il nautilo ha questa capacità e, di recente, ho potuto studiare il fenomeno su esemplari appena catturati di *Nautilus macomphalus* della Nuova Caledonia. In questa zona, almeno una conchiglia di nautilo su dieci presenta cicatrici di grosse fratture; simulando attacchi, sono stato in grado di misurare a quale velocità l'animale in questione poteva riempire la propria conchiglia.

Se un nautilo veniva costretto a portarsi in superficie, il liquido iniziava a rientrare piano piano praticamente in tutte le camere, entro la prima ora. Il tasso di riempimento era comunque molto basso: il massimo incremento di peso in 25 esemplari fu inferiore a 0,5 grammi all'ora. Anche la quantità totale di liquido che poteva penetrare nelle camere sembrava limitata. Nessuna camera singola veniva riempita almeno fino a metà e nessun nautilo (nemmeno quelli che avevano perso anche 25 grammi di conchiglia) aveva aumentato il proprio peso oltre i cinque grammi. Questo adattamento ai fini del galleggiamento compensativo sembrava essere estremamente circoscritto, almeno nel nautilo.

Che cosa possiamo dire in merito alle ammoniti? Anche molti dei loro fossili presentano fratture rimarginate sulla conchiglia. Per di più, le coste ispessite di quest'ultima sono orientate radialmente e pertanto le fratture presso la bocca della conchiglia, che è la parte più vulnerabile, avrebbero seguito un andamento parallelo a essa, con minima perdita di materiale. L'anatomia e la posizione del sifone delle ammoniti differiscono da quelle del nautilo, ma i miei studi indicano che la velocità di svuotamento e di riempimento delle camere del nautilo è in gran parte determinata dal rapporto fra area superficiale dell'epitelio del sifone all'interno di una qualsiasi camera e il volume della camera stessa; questi rapporti sono del tutto corrispondenti fra nautilo e ammoniti. Le prove non sono quindi decisive. Forse le ammoniti possedevano un sistema di riempimento simile, in quanto a capacità, a quello dei nautili; forse era molto più efficiente. È comunque chiaro che un simile sistema era indispensabile.

L'ultima e più pronunciata tendenza delle ammoniti del Cretaceo è verso l'eteromorfismo, condizione di quelle specie che presentano conchiglie per molti versi differenti dalla forma planispirale ancestrale. Qualche specie eteromorfa era apparsa per breve tempo fra le migliaia di specie con conchiglia planispirale del Triassico e del Giurassico. Tuttavia, a partire da 120 milioni di anni fa specie eteromorfe cominciarono a evolversi in grande quantità e, alla fine del Cretaceo, esse costituivano la maggioranza delle ammoniti, per numero sia di individui sia di specie.

È questa la tendenza più difficile da interpretare; le forme delle conchiglie delle specie eteromorfe variano dalle forme planispirali con spirale larga fino ad avvolgimenti bizzarri, che non sembrano aderire a uno schema qualsiasi. Alla fine del XIX secolo e durante il XX secolo, i biologi ipotizzarono che queste forme aberranti fossero gli ultimi rappresentanti, non adattabili, della lunga storia evolutiva degli ammonoidi, che preannunciavano l'estinzione delle ammoniti. Questo concetto di «tipolisi», o senescenza razziale, è stato accantonato e oggi la maggior parte dei ricercatori ritiene che le forme eteromorfe di conchiglie corrispondano ad adattamenti specializzati a una varietà di habitat.

Nella loro diversità di forme le specie eteromorfe rendono però difficile qualsiasi semplice generalizzazione sul significato adattativo delle loro conchiglie. Le prime forme eteromorfe del Cretaceo comprendevano sia ortoconi (cioè coni diritti) sia conchiglie planispirali con spirali larghe o forme anomale con un curioso uncino all'estremità. Negli strati del Cretaceo inferiore della California settentrionale, che sono forse la fonte più ricca di fossili di ammoniti di questo periodo (tra breve sommerse da un bacino artificiale), molte ammoniti eteromorfe hanno una conchiglia massiccia con pesanti ornamentazioni.

Negli stessi strati si trovano anche conchiglie planispirali con pesanti ornamentazioni. Le specie eteromorfe degli inizi del Cretaceo, come pure le ammoniti con conchiglia planispirale e con grosse ornamentazioni, hanno avuto probabilmente origine da specie di ammonoidi planispirali con conchiglia scarsamente idrodinamica. Jost Wiedmann dell'Università di Tübingen ha dimostrato, basandosi sulle linee di sutura, che le specie eteromorfe in generale sono derivate da specie planispirali con conchiglia non idrodinamica e spesso hanno dato origine a discendenti con conchiglia planispirale. La plasticità della tipologia delle conchiglie fa pensare che le prime ammoniti eteromorfe, al pari delle ammoniti planispirali ornate, dipendessero per la propria sopravvivenza più dall'ornamentazione difensiva che dalla forma della conchiglia.

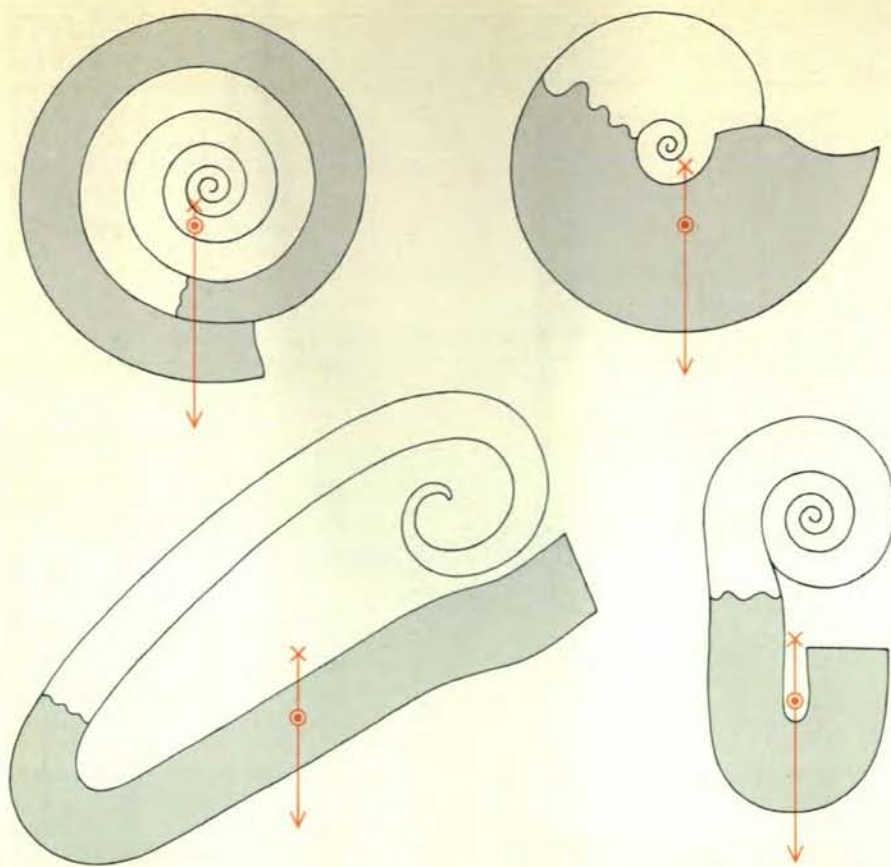
Alla fine della prima metà del Cretaceo, molte fra le più grosse specie eteromorfe si erano ormai estinte. Anche se forme giganti, come alcune specie del genere ortoconico *Baculites*, riapparvero nel tardo Cretaceo, la maggior parte delle specie eteromorfe di questo periodo era più piccola e ornata più delicatamente delle specie precedenti. Le forme eteromorfe del primo Cretaceo, come le loro antenate, probabilmente animali bentonici, mentre molte specie del tardo Cretaceo presentano adattamenti della conchiglia che fanno pensare che vivessero presso la superficie o a media profondità più che sul fondo marino.

Molte ammoniti eteromorfe del tardo Cretaceo hanno un avvolgimento elicoidale e si è ritenuto tradizionalmente che i loro modi di vita si conformassero a quelli dei molluschi ai quali maggiormente assomigliavano: i gasteropodi bentonici con

1 ORNAMENTAZIONE ASSENTE O MOLTO FINE — 2 DA FINE A MODERATA — 3 DA MODERATA A GROSSOLANA — 4 MOLTO GROSSOLANA

PERCENTUALE DEI VARI GENERI (0–80)

67 GENERI DEL PALEOZOICO — 211 GENERI DEL TRIASSICO — 150 GENERI DEL PRIMO GIURASSICO — 235 GENERI DEL MEDIO E TARDO GIURASSICO — 106 GENERI DEL PRIMO CRETACEO — 95 GENERI DEL TARDO CRETACEO

La presenza di coste robuste, un tipo di ornamentazione difensiva, era molto più comune nelle ammoniti del Cretaceo che negli ammonoidi più primitivi. L'ornamentazione è definita fine o grossolana in base al rapporto fra larghezza di una costa esterna e diametro della conchiglia.

La stabilità statica di gran parte delle conchiglie eteromorfe è ancor maggiore di quella delle conchiglie planispirali involute. Negli anni quaranta, Arthur E. Trueman dell'Università di Glasgow localizzò il baricentro e il centro di galleggiamento in una grande varietà di conchiglie di ammoniti. La posizione del baricentro dipende dalla distribuzione spaziale della cavità che contiene il corpo (*zona in grigio*), mentre la posizione del centro di galleggiamento dipende dalla distribuzione spaziale del fragmocono. La stabilità statica di molte conchiglie eteromorfe è tale che, con tutta probabilità, gli animali che abitavano in esse erano in grado di effettuare solo modeste variazioni di orientamento. L'orientamento normale di alcune avrebbe costretto il corpo dell'animale a rivolgersi verso la superficie anziché verso il fondale. Questo è stato il primo indizio comprovante che alcune specie eteromorfe, a differenza delle loro antenate, non erano bentoniche.

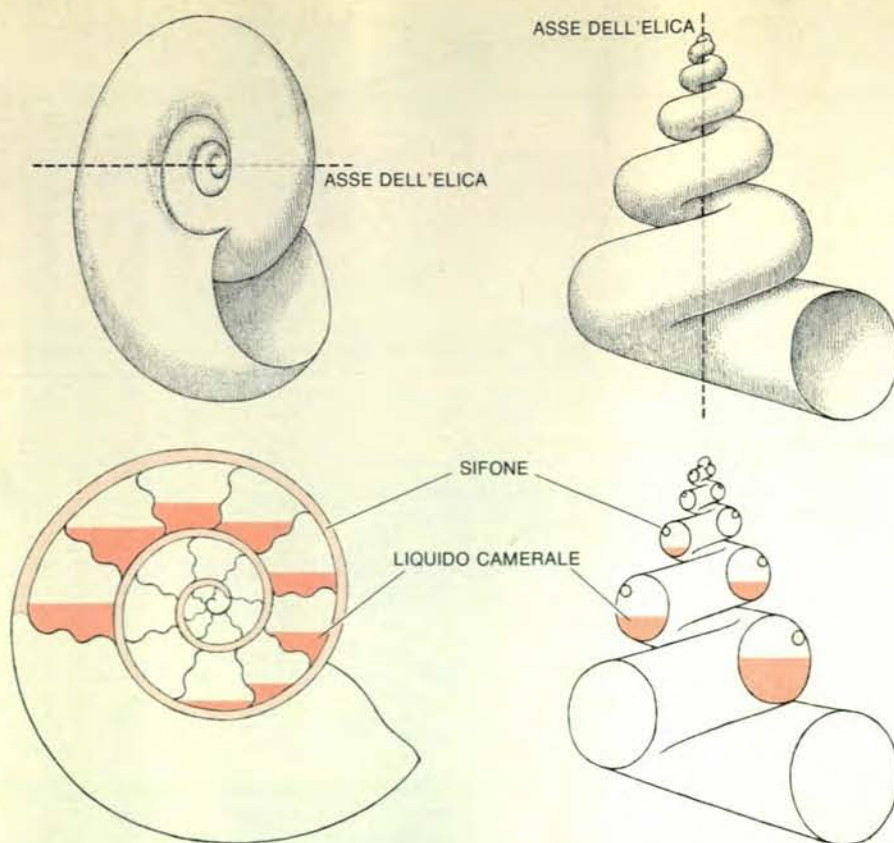
conchiglia elicoidale. La prima dimostrazione che ha portato alla reinterpretazione del loro modo di vita è stata data da Arthur E. Trueman, dell'Università di Glasgow. Agli inizi degli anni quaranta, Trueman ha calcolato i baricentri e i centri di galleggiamento di specie di ammoniti planispirali ed eteromorfe; i suoi calcoli mostrano che, di regola, le specie eteromorfe avevano una maggior stabilità statica delle specie planispirali, principalmente perché, con un minore avvolgimento della conchiglia, il baricentro nella camera contenente il corpo dell'animale di solito veniva separato dal centro di galleggiamento sito nel fragmocono. Molte specie eteromorfe avevano, effettivamente, una tale stabilità statica da metterle in grado di effettuare solo lievi variazioni nell'orientamento.

Se la conchiglia delle specie eteromorfe era avvolta in modo che la sua imboccatura (e quindi il corpo dell'animale) fosse rivolta verso il fondo marino, la sua stabilità non era incompatibile con la vita alle grandi profondità. Tuttavia, una frazione sempre più consistente delle specie eteromorfe che apparvero durante il Cretaceo presentava un uncino a forma di *U* all'estremità della camera contenente il corpo dell'animale. Di tanto in tanto i fragmoconi di queste conchiglie avevano forma a chiocciola e a volte si avvolgevano attorno a più di un asse oppure non seguivano per il loro avvolgimento alcuno schema. Comunque, in tutti i casi, l'estrema separazione fra fragmocono e camera contenente il corpo doveva rendere molto stabili queste conchiglie, costringendo inoltre il corpo dell'animale a rivolgersi verso l'alto, un orientamento estremamente scomodo per un organismo bentonico.

Un'ulteriore testimonianza ai fini della reinterpretazione dei modi di vita delle specie eteromorfe è stata fornita da Denton e da Gilpin-Brown, i quali hanno precisato che, nei cefalopodi attuali, il sifone è disposto in modo tale che il livello del liquido camerale scende al disotto di esso ancor prima che la camera sia completamente svuotata. Nel nautilo il sifone è in contatto diretto con il liquido camerale solo fino a quando la camera è semivuota; dopo che il liquido è stato «disinserito» (termine introdotto da Denton e Gilpin-Brown), la camera continua a drenarsi perché la pellicola, ossia il rivestimento membranoso del suo interno, è idrofila e agisce come uno stoppino facendo salire il liquido fino al sifone. Anche gli altri due generi attuali di cefalopodi a conchiglia concamerata, *Sepia* e *Spirula*, presentano adattamenti dello stesso tipo. Si ha l'impressione che le camere potrebbero essere svuotate in maniera più efficace se il sifone fosse orientato in modo da essere costantemente bagnato dal liquido camerale. Sono questi adattamenti accidentali oppure essi hanno uno scopo funzionale?

Il nautilo con conchiglia concamerata vive sul lato delle scogliere verso il mare aperto, laddove si verificano sensibili variazioni di profondità a intervalli orizzontali molto brevi. Bruce Carlson del Waikiki Aquarium di Hawaii, Mike Weekly del Seattle Aquarium e io stesso, mentre osservavamo - mediante telemetria con controllo a distanza - esemplari di *Nautilus belauensis* presso l'isola di Palau, nel Pacifico occidentale, abbiamo constatato che questi animali, mentre erano a caccia di cibo, seguivano i profili del fondale spostandosi spesso lungo notevoli distanze verticali in un breve periodo di tempo. Generalmente, i cinque esemplari sotto controllo si dirigevano nel corso della notte verso le acque basse, per ritornare in profondità durante il giorno, con dislivelli quotidiani di anche 330 metri. Anche *Spirula*, una piccola seppia, migra verticalmente, ma preferisce le medie profondità e non il fondo marino.

Il processo di svuotamento delle camere richiede quasi un mese in un piccolo nautilo e tre mesi negli esemplari più grossi. Per esso si ha un consumo di energia, anche se l'animale rimane a profondità costante, poiché la pressione idrostatica dell'acqua marina trasferita dal corpo dell'animale al sangue contenuto nel sifone tende a far rifluire liquido nella camera, vincendo il gradiente osmotico instaurato dagli enzimi presenti nelle pieghe dell'epitelio del sifone. Dato che la pressione idrostatica varia in funzione della profondità, anche il gradiente osmotico deve modificarsi in modo che il nautilo possa continuare a svuotare una camera o a impedire il riflusso in essa quando deve scendere a maggiore profondità. Ne consegue che qualsiasi mezzo inteso a «disinserire» il liquido camerale dal sifone costituirebbe un vantaggio. Quando un nautilo si sposta a una profondità diversa, se il sifone in una delle sue camere viene «disinserito», basta che la concentrazione salina cambi solo nel liquido assorbito dal tubo calcareo che circonda il sifone e nella pellicola con funzioni di stoppino all'interno della camera, invece che in tutto il volume del liquido camerale.

Il «disinserimento» del liquido camerale potrebbe aver avuto ancora maggiore importanza nelle ammoniti rispetto ai nautili; quando un nautilo ha completato lo sviluppo, tutto il liquido è stato praticamente pompato all'esterno delle sue camere. Recenti studi sulle ammoniti indicano, invece, che esse conservavano ancora una certa quantità di liquido anche dopo aver raggiunto la maturità. Richard

La posizione del sifone in alcune ammoniti eteromorfe può averle messe in grado di adattarsi più agevolmente alle variazioni di profondità, particolarmente nel corso della crescita, rispetto alle ammoniti con conchiglia planispirale. Sebbene l'anatomia del sifone di un'ammonite fosse leggermente diversa da quella di un nautilo, probabilmente il meccanismo utilizzato da quest'organo per svuotare le camere era in sostanza lo stesso. Gli enzimi presenti nelle cellule epiteliali del sifone determinano gradienti osmotici locali abbastanza rilevanti da permettere il drenaggio del liquido nonostante la pressione dell'acqua marina, che tende a forzarlo nella direzione opposta. Dato che la pressione varia in concomitanza con il livello di profondità il gradiente osmotico deve modificarsi, se l'animale deve continuare a svuotare le camere quando si sposta verso maggiori profondità. Se però il sifone rimane «disinserito» rispetto al liquido delle camere, è sufficiente che cambi il gradiente osmotico della membrana interna della conchiglia situata presso il sifone e che funge da stoppino. Nelle conchiglie planispirali, le camere formate per prime ruotano avvolgendosi verso l'alto quando viene aggiunto nuovo materiale presso la bocca della conchiglia; la posizione del sifone rispetto al liquido in qualsiasi camera, e conseguentemente il grado di «disinserimento» del sistema stesso, varia col procedere della crescita. In alcune conchiglie eteromorfe come la conchiglia torticonica (a forma di chiocciola) riprodotta a destra, in qualsiasi camera il sifone rimane in posizione costante nel corso della crescita e generalmente è «disinserito» dal liquido. Per effetto di questo adattamento, il dispendio energetico richiesto per svuotare le camere è minore.

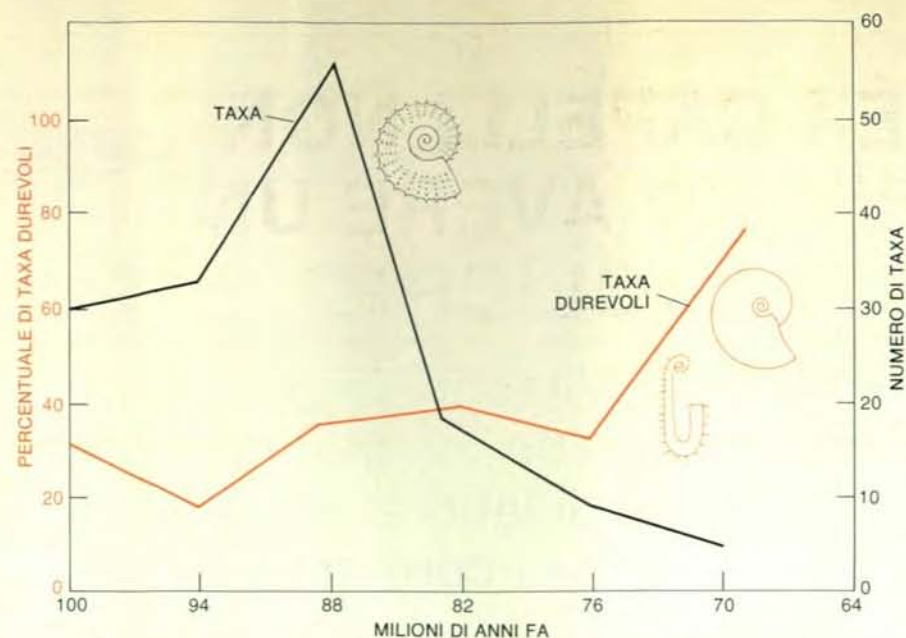
Reyment dell'Università di Uppsala è riuscito a calcolare il peso specifico di conchiglie di ammoniti a sviluppo completo partendo da fossili ben conservati e ha osservato che le conchiglie avevano una densità assai minore di quelle dei nautili. A meno chè la densità corporea delle ammoniti fosse maggiore di quella dei nautili, perché esse potessero avere un galleggiamento indifferente sarebbe stato necessario che il 10-20 per cento del loro fragmocono (in confronto all'1 o 2 per cento necessario per un nautilo adulto) fosse riempito di liquido.

Il «disinserimento» del liquido camerale nelle specie di ammoniti planispirali potrebbe però essere stato ancora meno efficiente che nei nautili della stessa tipologia. Quando un cefalopodo planispirale aggiunge nuovo materiale in corrispondenza della bocca della conchiglia, le camere formate per prime ruotano lentamente verso l'alto al disopra della camera di abitazione, che contiene il corpo dell'animale. Nel nautilo le posizioni di queste camere sono relativamente poco importanti, sia perché le camere sono vuote, sia perché il sifone è collocato centralmente. Nelle ammoniti, invece, le camere formate per prime erano probabilmente in parte ripiene di liquido e il sifone era disposto contro la parete esterna della conchiglia. In questo modo, il sifone all'interno delle camere poste alla sommità dell'avvolgimento sarebbe stato «disinserito» se una qualsiasi quantità di liquido fosse stata rimossa dalla camera, mentre la parte di sifone contenuta nelle camere alla base della conchiglia sarebbe rimasta collegata finché la camera conteneva del liquido. Il cambiamento di posizione del sifone conseguente all'accrescimento della conchiglia avrebbe resa ancora più difficoltosa per l'animale la conservazione del galleggiamento indifferente con il variare della profondità.

Uno dei principali vantaggi della conformazione delle conchiglie può essere stato quello che, mentre la conchiglia si accresceva, il sifone, all'interno di una qualsiasi camera, rimaneva in posizione costante rispetto al liquido camerale. Il miglior esempio di questo adattamento è offerto dalle specie di ammoniti appartenenti alla famiglia turrilitidi del medio Cretaceo e alla famiglia nostoceratidi del medio e tardo Cretaceo. In entrambe queste famiglie le conchiglie hanno un avvolgimento elicoidale: in una conchiglia planispirale l'asse di avvolgimento è orizzontale, mentre nelle conchiglie elicoidali esso è verticale. Di conseguenza, durante l'accrescimento della conchiglia, le camere seguono l'andamento della ruota panoramica di un Luna Park in una conchiglia planispirale e quello di una giostra in una conchiglia.

In una conchiglia elicoidale, il sifone è in posizione leggermente diversa rispetto al liquido camerale contenuto nei giri successivi della conchiglia, ma conserva questa posizione a mano a mano che la conchiglia si sviluppa; in una conchiglia planispirale esso non si comporta allo stesso modo. In molte specie delle due famiglie di ammoniti sopra citate, inoltre, il sifone si spostava verso l'alto seguendo la parete laterale della conchiglia, in modo che il «disinserimento» del liquido camerale si verificasse non appena aveva inizio l'evacuazione della camera. Questo spostamento, unitamente alla conservazione della posizione del sifone nel corso dell'accrescimento, stava a significare che la maggior parte delle camere del fragmocono non erano più collegate, pur conservando una certa quantità di liquido. Pertanto sembra che le specie eteromorfe, per conservare il galleggiamento indifferente, abbiano elaborato un sistema più efficiente, a spese di una ridotta idrodinamicità.

La grande stabilità delle conchiglie eteromorfe e gli adattamenti che queste presentano per disinserire il liquido camerale, hanno indotto alcuni ricercatori, come Andrew Packard dell'Università di Edimburgo e Herbert Klinger del Sud Africa a convenire con la mia supposizione che almeno alcune specie eteromorfe vivessero a media profondità invece che sul fondo oceanico. Molti cefalopodi attuali, come i membri del genere *Spirula* e della famiglia cranchidi, galleggiano come palloncini a media profondità, spostandosi verso le acque basse di notte e ritornando lentamente negli habitat più profondi durante il giorno. Può darsi che alcune specie di ammoniti eteromorfe siano state in grado di condurre questo genere di vita, evitando così gli habitat poco profondi della piattaforma continentale, nei quali vivevano le ammoniti del Cretaceo.

Tutti e tre i tipi di conchiglia (planispirale idrodinamico, planispirale con ornamentazione massiccia o corazzata, ed eteromorfo) sono comparsi in determina-

La riduzione della diversità, che precedette la completa estinzione delle ammoniti alla fine del Cretaceo, presenta una connotazione interessante: i 300 milioni di anni della storia degli ammonoidi sono caratterizzati da rapide evoluzioni ed estinzioni di molti taxa dalla vita breve. Durante il Cretaceo, però, la vita media dei generi di ammoniti andò aumentando in concomitanza con la subitanea riduzione numerica di tali generi. La riduzione della diversità è quasi interamente da imputare all'estinzione dei taxa caratterizzati da ricambio più rapido. Non è ancora chiaro per quale ragione si siano estinti i taxa più durevoli, alcuni dei quali erano molto affini ai nautiloidi attuali.

ti momenti della lunga storia degli ammonoidi. Ciò che ha distinto il Cretaceo è stato il subitaneo aumento di specie che presentavano questi tipi di conchiglia a svantaggio di quelle con conchiglie planispirali scarsamente idrodinamiche e con ornamentazioni non molto rilevanti, che prevalsero nella maggior parte della storia degli ammonoidi.

Durante le ultime fasi del Cretaceo (dal Campaniano al Maestrichtiano), il numero di specie di ammoniti subì una brusca contrazione. Questa riduzione della diversità presenta uno schema molto interessante. Con Philip W. Signor, dell'Università della California a Davis, ho analizzato le tendenze del tasso di comparsa e di quello di estinzione in famiglie, generi e specie di ammoniti del Giurassico e del Cretaceo. In questi due periodi geologici, il tempo medio di vita dei generi di ammoniti si aggirava sui sette milioni di anni mentre William Kennedy dell'Università di Oxford e William A. Cobban dell'US Geological Survey hanno calcolato che questo tempo medio, per le specie, era compreso fra uno e due milioni di anni. Tuttavia, tra questi taxa di breve durata vi sono alcuni generi e specie molto più durevoli: alcuni generi riuscirono a sopravvivere persino per 100 milioni di anni. Dai nostri studi risulta che il tempo medio di vita dei generi delle ammoniti aumentò verso la fine del Cretaceo; in altri termini, la riduzione della diversità fra ammoniti fu imputabile quasi interamente all'estinzione dei taxa a vita breve.

L'estinzione di questi taxa a vita breve e con conchiglie pesantemente ornate e non idrodinamiche spiega, in massima parte, la riduzione di diversità verificatasi alla fine del Cretaceo, e questo non dovrebbe sorprendere. Sopravvissero relativamente pochi taxa, ma dotati di una vita relativamente lunga; la maggior parte aveva conchiglie eteromorfe o idrodinamiche. Si potrebbe quindi facilmente obiettare che la grande riduzione di diversità non costituiva necessariamente la premessa per la completa estinzione delle ammoniti.

Non è ancora chiaro quale sia stata la sorte delle altre specie di ammoniti, poiché l'estinzione di questi animali si verificò nel periodo in cui i bacini oceanici poco profondi, il loro habitat preferito, si erano ridotti su scala mondiale. La maggior parte delle sezioni stratigrafiche comprendenti la transizione dal Cretaceo al Terziario riguardano pertanto sedimenti di mari profondi, in cui grossi fossili come le ammoniti sono assenti. Fino a questo momento sono state scoperte solo due sezioni stratigrafiche che registrano gli ultimi uno o due milioni di anni della storia delle ammoniti. In entrambe queste sezioni, negli strati del Maestrichtiano superiore, sono stati trovati fossili di circa 10 specie di ammoniti, e circa la metà appartiene a generi con conchiglie compresse e idrodinamiche, come *Sphenodiscus* e *Pachydiscus*, mentre l'altra metà appartiene a generi eteromorfi come *Scaphites*, *Baculites* e *Diplomoceras*. Purtroppo, le testimonianze di questi strati sono per altri versi contraddittorie.

Nella sezione Stevns Klint, in Danimarca, Tove Birkelund dell'Università di Copenhagen ha trovato ammoniti di varie dimensioni immediatamente al disotto degli strati contenenti concentrazioni più alte del normale di iridio, uno degli elementi del gruppo del platino che gli Alvarez e i loro collaboratori pensano derivino dall'impatto di un meteorite. Di recente, invece, Wiedmann e io abbiamo potuto raccogliere a Zumaya, in Spagna, un gran numero di ammoniti nella sezione corrispondente al tardo Cretaceo; come nella sezione danese, le ammoniti sono distribuite in abbondanza in tutti gli strati del Cretaceo e centinaia di fossili possono essere raccolti dai piani di stratificazione bene esposti.

A circa 15 metri al disotto del deposito anomalo di iridio nella sezione di Zumaya, il numero delle ammoniti gradualmente decresce e l'ammonite più recente, in qualsiasi prelievo effettuato, proviene da 10 metri al disotto del limite di demarcazione, anche se, immediatamente al di sotto e al di sopra di esso si trovano altri grossi fossili, ad esempio di echinoidi e di bivalvi. La prova è negativa e potrebbe essere ribaltata dal rinvenimento di un solo nuovo campione di ammonite. Ma, anche se venisse trovato un esemplare al disotto di tale limite, non muterebbe il fatto che le ammoniti erano virtualmente scomparse da Zumaya molto prima del supposto impatto di un meteorite.

Qualsiasi ipotesi sulla sorte delle ammoniti deve tener conto di una questione fondamentale. Per quale ragione, alla fine del Cretaceo, persino le specie di ammoniti che erano sopravvissute a lungo si estinsero mentre i nautiloidi sopravvissero? Forse non si potrà mai fornire una risposta decisiva. Posso solo presumere che la strategia riproduttiva o forse qualche aspetto dell'ecologia degli adulti abbia salvato i nautiloidi. Il plancton, l'insieme di piante e animali microscopici galleggianti nel mare, ha subìto alla fine del Cretaceo, quando scomparve il 90 per cento delle specie planctoniche, il maggior tasso di estinzione rispetto a qualsiasi altro gruppo di organismi marini. Le giovani ammoniti nate da piccole uova, con conchiglie che non superavano il millimetro di diametro, possono aver trascorso i loro primi giorni o settimane mescolate col plancton. In base alla forma della conchiglia, sembra invece che, dal Mesozoico in avanti, i giovani nautiloidi avessero dimensioni molto maggiori (da cinque a 25 millimetri). Probabilmente non si mescolavano al plancton, assumendo subito le abitudini alimentari degli adulti, cioè andando a cercare il cibo in profondità, vicino ai fondali. Può darsi, quindi, che le ammoniti siano state coinvolte nel collasso degli ecosistemi planctonici, sia in giovane età sia da adulte allorché si nutrivano a un livello della catena alimentare inferiore a quello dei nautiloidi.

Quale sia stata la ragione di questo fatto, i nautiloidi superarono il periodo critico del tardo Cretaceo e persino si irradiarono, anche se in misura modesta, nel Terziario. La scarsa diversità dei nautiloidi attuali indica, tuttavia, che il tempo delle conchiglie concamerate, per quanto ben strutturate, è ormai trascorso.

(RI)CREAZIONI AL CALCOLATORE

di Brian Hayes

Dove si introduce una nuova rubrica dedicata ai piaceri del calcolo

Calculemus
—Gottfried Wilhelm von Leibniz

Alcuni dei migliori giochi col microcalcolatore sono ora dotati di una funzione detta TBIC. Le lettere stanno per *The boss is coming* (arriva il capo) e quando si preme il tasto assegnato alla funzione lo schermo diventa immediatamente scuro e silenzioso. Sono qui rappresentati, mi sembra, i due poli della risposta pubblica alla recente proliferazione di calcolatori a basso costo. Da una parte il calcolatore è una macchina per il mondo degli affari, uno strumento capitalistico; dall'altra è un mezzo di intrattenimento così frivolo da dover essere nascosto, come i fumetti.

Non intendo sminuire nessuna delle due applicazioni pratiche dei calcolatori: gli affari e l'industria o quel genere di giochi il cui scopo principale è mettere alla prova i riflessi del giocatore. Il calcolo a fini utilitaristici è indubbiamente importante e, in quanto ai videogiochi, la loro costruzione può essere tra le espressioni più alte dell'arte della programmazione. Va osservato, però, che né l'uno né l'altro di questi usi del calcolatore implicano in profondità la questione di ciò che un calcolatore è e di ciò che esso può fare.

Esiste un vasto territorio compreso tra la programmazione per affari e i videogiochi, tra VisiCalc e Space Invaders. In questo territorio sono comprese le applicazioni del calcolatore a tutte le arti e scienze e, forse più ovviamente, alla matematica. È compreso l'uso del calcolatore per simulare aspetti del mondo naturale e delle società umane. Sono compresi, inoltre, molti impieghi che, propriamente parlando, non sono affatto «usi» del calcolatore ma servono piuttosto a centrare l'attenzione sul calcolatore stesso e sulla natura del calcolo meccanizzato. Intento della nostra rubrica sarà esplorare questo campo tra la pratica più austera e il mero divertimento. Saranno quanto mai graditi gli interventi dei lettori e, dato che il pensiero rimane di gran lunga più efficace del calcolatore, coloro che non hanno possibilità di accesso a un calcolatore saranno appena di poco svantaggiati.

Anche gli strumenti dell'uomo d'affari si possono a volte applicare a problemi relativi alla teoria e alla pratica del calcolo meccanico. Considererò qui alcune questioni sollevate da applicazioni non convenzionali dei programmi detti *spreadsheet*, cioè tabelloni o fogli elettronici.

Un tabellone di carta è un grande foglio diviso in righe e colonne che potrebbe essere impiegato per analizzare il bilancio di un'azienda. A ogni reparto potrebbe essere attribuita una colonna, e una riga a ogni conto di ricavo o di perdita. In altre colonne e righe potrebbero essere inseriti i totali e le percentuali per ogni reparto e per ogni conto.

Il tabellone elettronico riproduce questa struttura sullo schermo di un tubo catodico, ma con alcune rilevanti differenze. Sulla carta, una casella (definita come l'intersezione di una colonna e una riga) può portare o una dicitura, come il nome di un reparto, o un numero. In un tabellone elettronico a una casella può essere assegnata anche una formula matematica. Così la casella alla fine di una riga potrebbe contenere una formula che richiede di sommare i valori immessi in tutte le altre caselle della riga. Sullo schermo appare il numero risultante dalla valutazione della formula, in questo caso il totale, ma il contenuto di fondo della casella è la formula stessa e non il numero. Se si modifica uno degli altri ingressi della riga, il totale viene automaticamente ricalcolato.

Il primo dei programmi di tabellone elettronico fu VisiCalc, elaborato nel 1978 da Daniel Bricklin, allora studente alla Harvard University School of Business, da Robert Frankston e Dan Fylstra. Si dice che ne siano state vendute più copie di qualsiasi altro programma per calcolatore. In seguito sono stati introdotti decine di altri programmi che operano su principi analoghi e lo stesso VisiCalc è stato più volte revisionato. La maggior parte degli esperimenti qui descritti sono stati effettuati con due più recenti programmi di tabellone: 1-2-3, ideato da Mitchell Kapor della Lotus Development Corporation di Cambridge, Massachusetts, e Multiplan, un prodotto della Microsoft Corporation di Bellevue, Washington. Nella maggior parte dei casi, andrebbero altrettanto bene altri programmi a tabellone.

Sebbene concepito per l'analisi finanziaria, il tabellone elettronico è capace di molto di più. È una matrice bidimensionale di caselle in cui il valore di ogni casella può essere fatto dipendere da qualsiasi altra casella o gruppo di caselle. È sorprendente vedere quanto della struttura matematica del mondo possa essere inserito in tale formato; il tabellone elettronico risulta un contesto di grande generalità per la descrizione di relazioni logiche e matematiche.

Un semplice esempio può contribuire a una migliore comprensione di come si utilizza un tabellone e di quali siano le sue potenzialità. Ogni casella è identificata dalle sue coordinate; nella maggior parte dei programmi le colonne sono individuate da lettere e le righe da numeri, partendo dall'angolo a sinistra in alto. Supponiamo che tanto alla casella A1 quanto alla casella A2 sia assegnato un valore numerico 1. Si immette poi una formula nella casella A3: il valore della posizione A3 è dato dalla somma del valore della casella immediatamente sopra essa e del valore della casella sopra quest'ultima. In altre parole, A3 è uguale al contenuto di A2 più il contenuto di A1 ed esibisce il valore 2.

Quello che è stato compiuto finora è banale: si tratta di uno schema molto elaboratoperesprimerelarelazione 1 + 1 = 2. Ora, però, è possibile riprodurre la formula di A3 in molte altre caselle. (L'esatta procedura per questa riproduzione varia da programma a programma, ma tutti i programmi prevedono questa possibilità.) Supponiamo che la formula di A3 sia copiata nelle caselle da A4 ad A10. Ognuna di quelle caselle conterrà un valore uguale alla somma dei valori delle due caselle superiori. Si noti che le formule sono tutte identiche ma, dato che sono applicate a valori diversi, i risultati non lo sono. I numeri esibiti, leggendo dall'alto verso il basso, sono 1, 1, 2, 3, 5, 8, 13, 21, 34 e 55.

Vi sono molti modi per generare la serie di Fibonacci con un calcolatore e la maggior parte di essi fanno un uso molto più efficiente delle risorse della macchina. C'è però qualcosa di caratteristico nella strategia del tabellone: non è algoritmica. In quasi tutti i linguaggi di programmazione un compito o la soluzione di un problema sono definiti in termini di un algoritmo, cioè di una serie di istruzioni esplicite che debbono essere eseguite in una sequenza ben determinata. Un algoritmo è un po' come una ricetta: potrebbe iniziare con «Mescolare farina, lievito e acqua, lasciar montare e infine cuocere». Se effettuassimo le stesse operazioni in un altro ordine avremmo un risultato molto differente. Il tabellone non ha questa caratteristica di ordinamento temporale. Nelle caselle non va inserita una sequenza di passi che porta dal problema alla soluzione, ma una struttura statica che cerca di racchiudere tutta insieme l'intera procedura. È una descrizione, anziché una ricetta: afferma che il pane è fatto di farina, lievito e acqua mescolati, fatti montare e cotti.

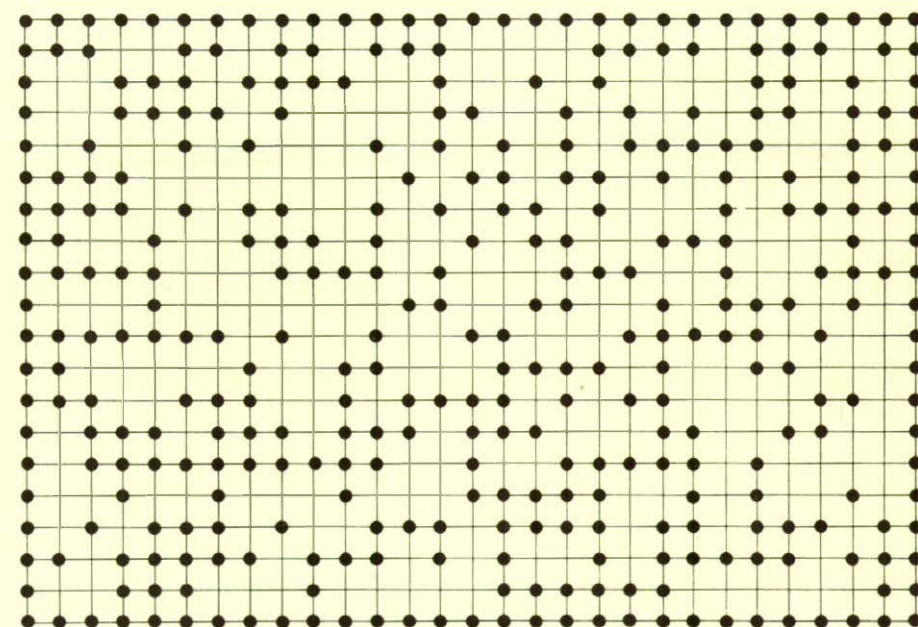
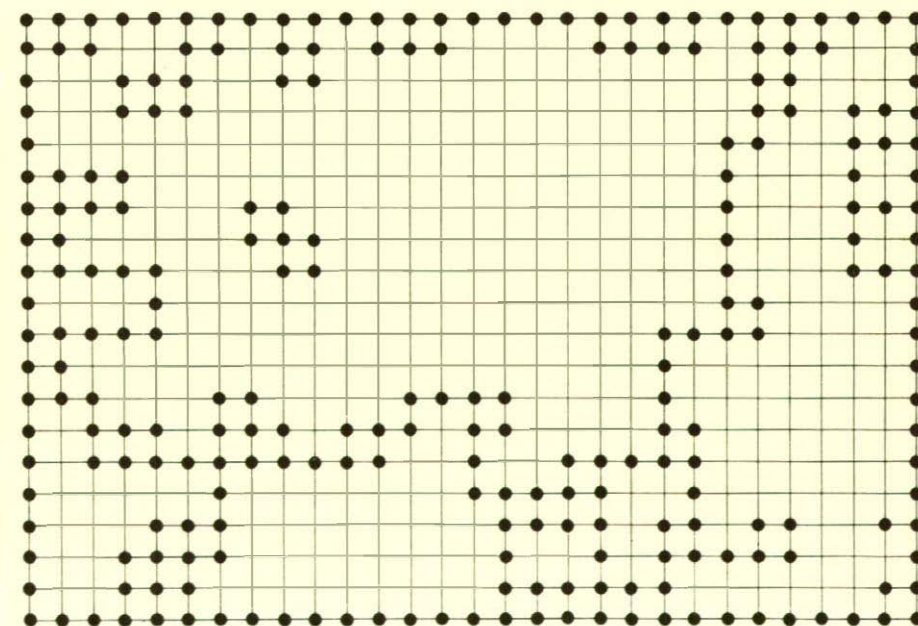
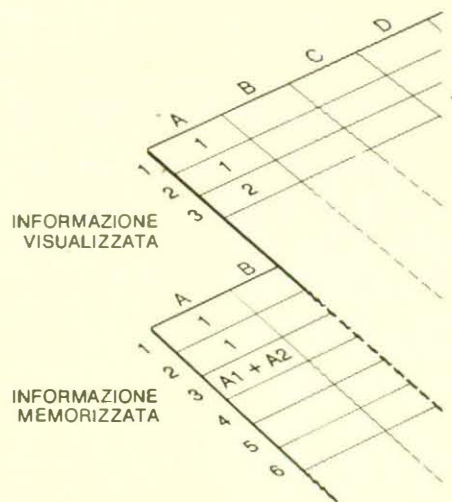
Possiamo chiarire meglio la distinzione tra un algoritmo e una descrizione statica con un altro esempio. Consideriamo il procedimento per moltiplicare due matrici di numeri, ciascuna delle quali abbia tre colonne e tre righe. L'algoritmo standard inizia con le istruzioni di moltiplicare ogni elemento della prima colonna della prima matrice per ogni elemento della prima colonna della seconda matrice, di sommare i tre risultati e immagazzinare la somma come primo elemento della matrice prodotto. Le stesse istruzioni sono poi ripetute per le altre otto combinazioni di righe e colonne. In un tabellone il problema è posto in altra forma e può sfruttare l'analogia matematica tra una matrice matematica e una schiera di caselle. Invece di scrivere una sequenza di istruzioni, si definisce semplicemente la matrice prodotto, ponendo ogni casella uguale a una formula che rappresenta l'opportuna combinazione di colonne e righe. Una volta immesse le formule, queste vengono risolte «tutte in un colpo» e appare l'intera matrice prodotto.

A un livello più profondo, naturalmente, un calcolatore che lavora sotto la direzione di un programma di tabellone sta in realtà eseguendo un algoritmo. Un calcolatore che abbia una sola unità centrale di elaborazione può fare una cosa sola alla volta e quindi le caselle sono valutate secondo una certa sequenza. Chi usa il programma, però, di solito non tiene conto della sequenza e in effetti spesso ne è del tutto all'oscuro: non ha quindi bisogno di pensare in termini di algoritmi.

Non è certo mia intenzione suggerire che la modalità di pensiero non algoritmica sia in qualche modo migliore di quella algoritmica: qualcuno può anche preferirla, ma è in gran parte questione di gusto. Quando una procedura diviene molto complessa, ci sono buone ragioni per raccomandare l'algoritmo, più facilmente divisibile in pezzi maneggevoli. Per risolvere il problema tutto in una volta bisogna capirlo tutto in una volta. Sembra comunque probabile che ci siano certi problemi o classi di problemi che si prestano naturalmente a una formulazione non algoritmica.

Un campo in cui la disposizione bidimensionale del tabellone risulta appropriata è la costruzione di sistemi di automi cellulari. Lo studio di questi sistemi venne iniziato negli anni cinquanta da John von Neumann e Stanislaw Ulam, che a quel tempo si interessavano soprattutto di modelli autoriproducentesi. Le regole fissate da von Neumann e Ulam richiedono uno «spazio cellulare uniforme» in cui ogni cellula rappresenta un automa, cioè una macchina che ha solo un numero finito di stati possibili. Lo spazio è uniforme nel senso che le leggi che governano lo stato degli automi sono le stesse per tutte le cellule. Un'ulteriore limitazione è che lo stato di una cellula può essere influenzato solo dalla sua storia e da quella dei suoi immediati vicini.

Le condizioni che definiscono un sistema di automi cellulari possono essere facilmente soddisfatte da un programma di tabellone. In linea di principio, il numero di stati disponibili a una cellula (che sarà rappresentata da una casella del tabellone) è enorme (forse 10^{100}), ma è chiaramente finito e può essere ridotto a un piccolo numero se sembra opportuno; per esempio, a una cellula può essere assegnata una formula che può assumere solo due possibili valori, quali 0 e 1. La richiesta di uniformità aggiunge un'interessante limitazione: essa implica che ogni cellula in cui è scritta una formula contiene sempre precisamente la stessa formula. (C'è più di un modo per decidere se due formule sono uguali. Supponiamo che una formula nella casella A1 si riferisca alla casella B1 immediatamente sotto di essa. Una formula in A2 dovrebbe essere considerata identica se anch'essa si riferisce a B1, con identità di «indirizzo assoluto», oppure se si riferisce a B2, con mantenimento della relazione geometrica. La seconda interpretazione è di solito più utile e sembra più consona alle idee di von Neumann e Ulam, ma entrambi gli schemi sono accettabili se applicati in modo coerente. I sottoprogrammi del tabellone per copiare i contenuti delle caselle forniscono un semplice test operativo di uniformità. Una schiera di caselle può essere considerata uniforme se, una volta inserita in una casella, una formula può poi essere copiata dal programma in tutte le altre.)

Von Neumann riuscì a dimostrare che

Reticolo generato da un programma di tabellone elettronico che simula la percolazione

«Assottigliando» il reticolo si rendono più evidenti i percorsi continui

Struttura di un tabellone elettronico

esiste una configurazione autoreplicantesi di cellule e ci riuscì seguendo una via estremamente complessa, mostrando che esiste un costruttore universale che può creare qualsiasi modello e deve essere quindi in grado di costruire il propriostesso modello. La dimostrazione richiede 200 000 cellule con 29 stati possibili. Per quanto ne so io, il modello non è mai stato costruito realmente, né manualmente né con l'aiuto di un calcolatore. È comunque pensabile che rientri nelle possibilità dei programmi di tabellone più potenti.

Un sistema molto più semplice in grado di autoreplicarsi fu escogitato nel 1960 da Edward Fredkin, del Massachusetts Institute of Technology. Ogni cellula ha solo due stati possibili, viva e morta, che possono essere rappresentati dai numeri 1 e 0. Lo stato di una cellula della generazione successiva è determinato dallo stato presente dei suoi quattro adiacenti ortogonali, cioè delle quattro cellule immediatamente a nord, est, sud e ovest. Se il numero dei vicini vivi è pari (0, 2 o 4), la casella muore o rimane morta. Invece, una casella con un numero dispari di vicini vivi (1 o 3) vive.

È semplicissimo esprimere questa regola in una formula per tabellone, soprattutto con programmi che includono una funzione per l'aritmetica modulare. Nel caso della casella B2 la formula è semplicemente (B1 + C2 + B3 + A2) modulo 2. L'effetto della formula è sommare i valori delle quattro caselle adiacenti, dividere per 2 e conservare solo il resto, che è necessariamente 0 o 1. Non rimane poi che copiare la formula (in modo che i riferimenti delle caselle preservino le stesse relazioni geometriche) in tutte le caselle di una regione del tabellone. In realtà c'è un'altra sottigliezza nella costruzione del sistema: sono necessarie due copie dello spazio cellulare. Una copia rappresenta l'attuale generazione e una preserva lo stato della generazione precedente. Lo stato presente di una cellula si fonda sul numero degli adiacenti vivi nella generazione precedente.

Quando il tabellone è disposto secondo le regole di Fredkin e viene fornito un modello iniziale, ogni ciclo di ricalcolo produce un nuovo modello. Dopo alcuni cicli, appaiono quattro copie della configurazione originale. In seguito le stesse copie sono copiate quattro volte e il modello iniziale è stato quindi riprodotto 16 volte. Il numero di cicli necessario per la riproduzione dipende dalla complessità del modello iniziale; nel caso più semplice (una sola casella viva) i quattro discendenti appaiono immediatamente.

Può risultare affascinante osservare il progredire di una colonia in crescita. La simmetria si mantiene tutte le volte e alcuni dei modelli hanno una sorprendente forma stellata. C'è un ritmo nel processo: il perimetro dell'area occupata si espande in modo continuo, ma l'interno si riempie periodicamente di un denso strato di cellule e poi si svuota di nuovo.

Sicuramente il più noto fra i sistemi di automi cellulari è il gioco «Vita», inventato da John Horton Conway dell'Università di Cambridge e fatto conoscere al mondo da Martin Gardner nella sua rubrica «Giochi matematici» (si veda «Le Scienze» n. 33, maggio 1971, e n. 35, luglio 1971). Il gioco, proprio come i suoi automi in continua moltiplicazione, si è ormai esteso a quasi tutti i sistemi di calcolatori e linguaggi di programmazione. C'è una buona ragione: ricompensa largamente l'attenzione prestatagli, sia essa un'attenzione da spettatore casuale oppure sia un'analisi approfondita.

Nel gioco Vita di Conway, le regole non sono definite per assicurare la replica di un modello ma per massimizzare la varietà o minimizzare la prevedibilità. Anche in questo caso ogni cellula ha due stati possibili, ma il suo stato è influenzato non più da quattro, ma dalle otto cellule più vicine, comprese quelle che hanno solo un vertice in comune con la cellula di partenza. Se una cellula è viva, continuerà a vivere solo se ha due o tre vicini vivi. Se ne ha di meno si dice che muore di solitudine e se ne ha di più muore per sovraffollamento. Per una cellula non viva è possibile nascere solo se ha esattamente tre vicini vivi.

Una specificazione algoritmica di questo procedimento tende a essere molto ripetitiva: bisogna esaminare una data cellula, contare i suoi vicini, decidere se deve vivere o morire, poi passare a una cellula successiva e poi a una successiva ancora finché tutte sono state verificate. La ripetizione di solito in un programma si ottiene con quella particolare struttura chiamata *loop* o ciclo, che in questo caso viene eseguita una volta per ogni cellula. Quando si codifica il gioco Vita in un tabellone, la ripetizione c'è ancora ma è spaziale piuttosto che cronologica: si immette la stessa formula in ogni casella di un'ampia configurazione.

Ci sono molti modi di scrivere una formula per valutare lo stato di una cellula nel gioco Vita. Il migliore che io abbia visto (nel senso di quello che procede più velocemente) è stato escogitato da Ezra Gotthcil della Lotus. Il procedimento base consiste nel moltiplicare il valore della cellula in esame per 9, ottenendo un risultato pari a 0 o 9, e poi sommare i valori delle otto cellule circostanti. Il risultato è confrontato con una tabellina che dà il nuovo stato della cellula per tutti i possibili valori della somma (vale a dire i valori compresi tra 0 e 17).

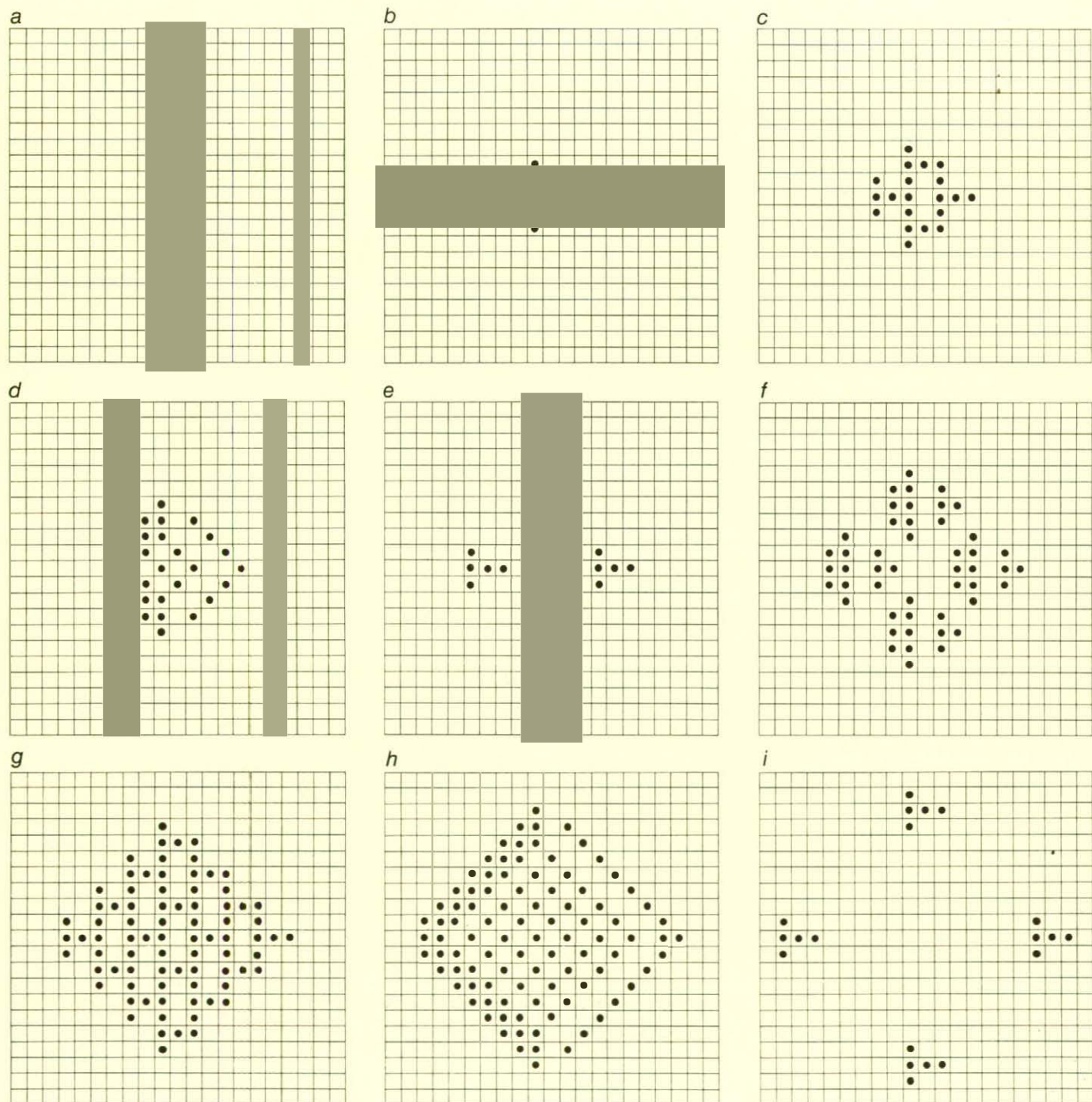
Idealmente Vita andrebbe giocato su una matrice cellulare di estensione infinita. Una delle attrattive del gioco è che certe piccole configurazioni iniziali si sviluppano, dopo appena poche generazioni, in magnifici fiori simmetrici; altre configurazioni emettono proiettili compatti che scivolano via a distanze infinite. L'evoluzione di un organismo si modifica, quando esso supera i confini del mondo. Una matrice infinita è in ogni modo impossibile e quando si lavora con un tabellone i limiti pratici sono, in realtà, piuttosto stretti e dipendono dalle possibilità del programma stesso, dalla capacità di memoria del calcolatore e dalla propria pazienza. (Il tempo necessario per creare una nuova generazione è, grosso modo, proporzionale al numero di cellule coinvolte.) Una strategia per creare una matrice che non abbia confini, pur essendo su un'area finita, è quella di definire adiacenti le cellule disposte lungo bordi opposti; si modifica così la topologia del tabellone. Unendo due bordi in tal modo si genera un cilindro o, se al foglio si fa compiere una torsione, un nastro di Möbius. L'unione dei quattro bordi a due a due dà luogo a un toro.

Il modello di Ising (si veda l'articolo *Problemi fisici con più scale di grandezza* di Kenneth G. Wilson, in «Le Scienze», n. 134, ottobre 1979) è la simulazione di un sistema fisico che, superficialmente, assomiglia a certi automi cellulari, benché la sua interpretazione sia molto differente. Il modello, che è stato introdotto dai fisici tedeschi Wilhelm Lenz ed Ernest Ising negli anni venti, può rappresentare molti fenomeni fisici, ma viene per lo più usato per la descrizione dei materiali ferromagnetici. Ogni posizione in un reticolo rappresenta il momento angolare di spin e quindi il momento magnetico di un atomo. Ogni spin ha modulo fisso, ma l'asse di spin può essere orientato «verso l'alto» o «verso il basso». Quando tutti gli spin hanno la medesima orientazione, il materiale è completamente magnetizzato; quando invece la loro orientazione è casuale, la magnetizzazione è nulla.

La realizzazione del modello di Ising con un tabellone elettronico è un po' più complicata di quella dell'automa replicantesi di Fredkin o del gioco Vita di Conway. Anche qui lo stato di un dato spin è influenzato dalle caselle più vicine, e precisamente dalle quattro ortogonali. Nel modello di Ising, tuttavia, c'è un elemento di casualità che rappresenta l'effetto della temperatura diversa da zero. Se gli spin delle caselle vicine sono tutti orientati verso l'alto, anche lo spin di quella data casella avrà la tendenza a

orientarsi verso l'alto, ma non è sicuro che ciò avvenga; la probabilità è inversamente proporzionale alla temperatura.

Alcuni esperimenti fatti da me con un modello di Ising su tabellone elettronico hanno dato risultati discordanti. Le proprietà del modello bidimensionale sono note esattamente dal 1944, quando Lars Onsager della Yale University ha risolto il sistema in modo analitico (anziché con una simulazione numerica). Quando la temperatura viene abbassata fino a un valore limite (la temperatura di Curie), gli spin cominciano a fluttuare selvaggiamente e poi si magnetizzano completamente. Non ho assistito a tale transizione di fase, ma la cosa non mi sorprende. Il modello di Ising richiede macchine con enormi potenze di calcolo. Per ottenere risultati accurati è necessario un reticolo molto grande ed è necessario esaminare tutte le possibili configurazioni degli spin, il che può richiedere molte ore anche con un programma efficiente e un elaboratore molto veloce. All'Istituto di fisica teorica dell'Università della California a Santa Barbara è stato costruito un calcolatore apposito per il modello di Ising che calcola 25 milioni di spin al secondo. Con il modello realizzato su tabellone elettronico, gli spin esaminati al secondo sono 25.

Anche se gli eventi interessanti che si verificano attorno alla temperatura di Curie non si possono osservare, il modello di Ising su tabellone sembra simulare bene altre proprietà dei materiali magnetici. A temperatura elevata la disposizione degli spin non evidenzia nessuna struttura, come ci si potrebbe aspettare. A bassa temperatura la magnetizzazione del reticolo è ovvia. Si creano spontaneamente ampi gruppi irregolari di spin allineati, mentre gruppi di polarità opposta sembrano scontrarsi ai loro confini. Sorprendente (almeno per me) è stata la comparsa di una fase antiferromagnetica in cui tutti gli spin hanno orientazione alternata (uno verso l'alto, uno verso il basso ecc.).

Sistema di automi cellulari ideati da Edward Fredkin in cui ciascuna struttura riproduce se stessa

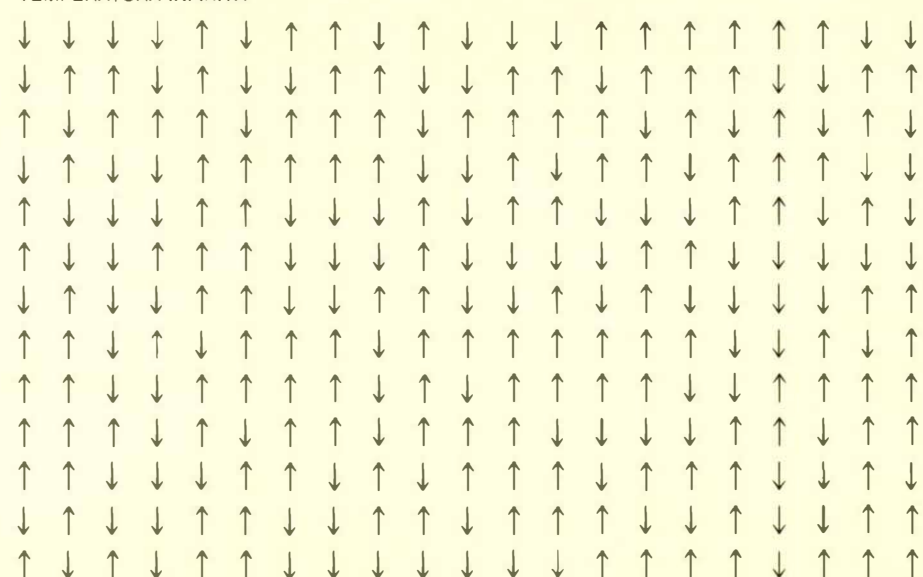
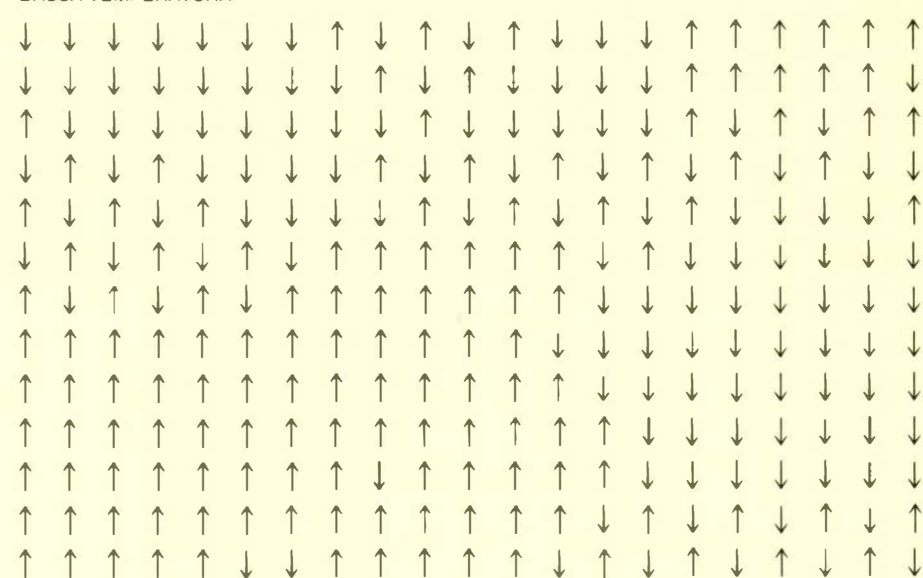
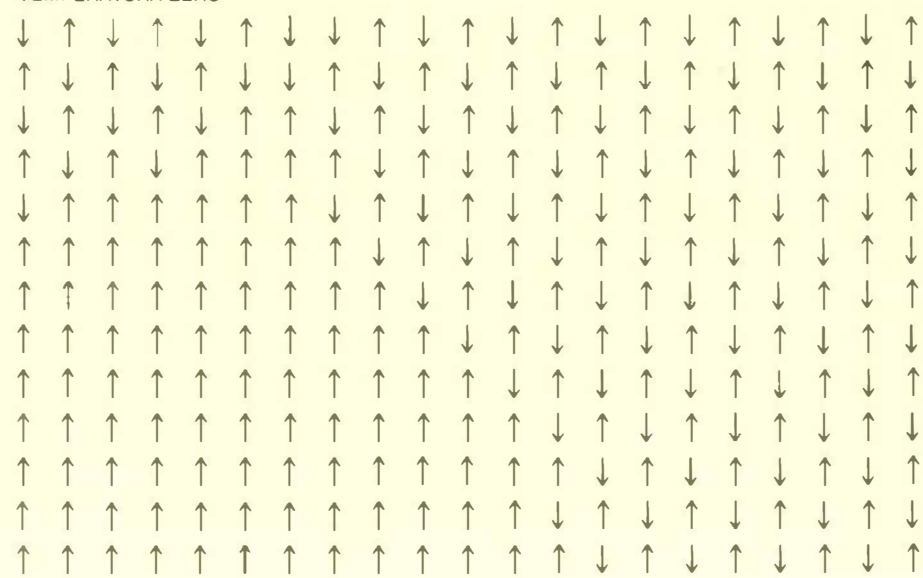
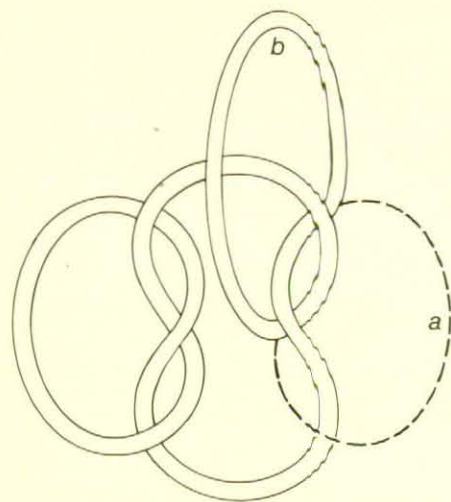
Allo zero assoluto la fase antiferromagnetica sembra la configurazione più stabile, ma ciò potrebbe riflettere un mio errore nella costruzione del modello.

Un reticolo piano permette di trattare anche altri sistemi fisici. Un fenomeno che mi interessa è la percolazione, che descrive non solo il modo in cui si fa il caffè, ma anche la struttura di alcuni polimeri, la conducibilità delle leghe, l'efficienza delle reti telefoniche, la propagazione degli incendi nei boschi e la diffusione delle malattie infettive. In un semplice modello della percolazione le caselle del reticolo potrebbero rappresentare possibili posizioni di atomi di rame in una matrice isolante. In prima approssimazione, la probabilità che una posizione sia occupata è proporzionale alla concentrazione di rame. Ciò che interessa è la probabilità di formare una catena ininterrotta di atomi di rame in qualche dominio del reticolo; è questa probabilità che determina la conducibilità del materiale.

Anche un modello della percolazione su tabellone elettronico comporta qualche forma di casualità. Il sistema più semplice consiste nel servirsi di uno spazio cellulare uniforme, in cui tutte le posizioni hanno la stessa probabilità di essere occupate e sono tutte indipendenti fra loro. Il risultato è una disposizione casuale di caselle libere e occupate. È allora necessario determinare se esiste o meno un percorso continuo attraverso lo spazio. Una soluzione corretta sarebbe quella di eseguire una ricerca sistematica di ogni percorso potenziale, ma non mi è chiaro come si possa fare senza ricorrere ad algoritmi. Una tecnica più brutale, ma ugualmente utile, è quella di includere nella formula per ogni casella una funzione che elimini ogni atomo con meno di due atomi adiacenti. Dato che ogni atomo che fa parte di una catena deve avere almeno due atomi adiacenti, la catena non viene modificata da questa procedura, ma ogni volta che il tabellone viene ricalcolato si eliminano le strade senza uscita e i gruppi rimasti isolati.

Un programma di tabellone elettronico non è certamente il mezzo ideale per rappresentare ognuno di questi modelli matematici o fisici. Per un lavoro serio, ciascun modello andrebbe realizzato con un programma «dedicato». Nel caso del gioco Vita ho messo a confronto varie versioni su tabellone con un programma (un algoritmo!) scritto nel linguaggio nativo di un microelaboratore. Per configurazioni della stessa dimensione, il programma in linguaggio macchina è quasi 100 volte più veloce del tabellone più veloce. La ragione non è difficile da trovare: anche se una casella non può mai avere un valore diverso da 0 o 1, il programma di tabellone calcola il valore fino alla quindicesima cifra decimale.

Se non può vantare la dote dell'efficienza, il tabellone ha in compenso la virtù della versatilità. Scrivere un programma in linguaggio macchina per giocare a Vita è ben più che uno svago pomeridiano. Inoltre, il programma non può fare nien-

TEMPERATURA INFINITA

BASSA TEMPERATURA

TEMPERATURA ZERO

Versione su tabellone del modello bidimensionale di Ising del ferromagnetismo

Come si dimostra che un nodo piano è alterno

t'altro, mentre la semplice matrice di caselle collegate in un tabellone costituisce un mezzo per la soluzione di problemi di impressionante generalità. Le possibilità dei programmi vanno ben al di là di quanto delineato in precedenza. Risulta evidente che si può generare qualsiasi serie di numeri i cui termini siano definiti da funzioni algebriche o trigonometriche. Si può costruire un crivello per individuare i numeri primi a partire da una breve formula ripetuta centinaia di volte. Si può rappresentare un campo fisico prendendo l'indirizzo di ciascuna casella come le sue coordinate in uno spazio bidimensionale. Come dice la pubblicità: «L'unico limite è la vostra immaginazione».

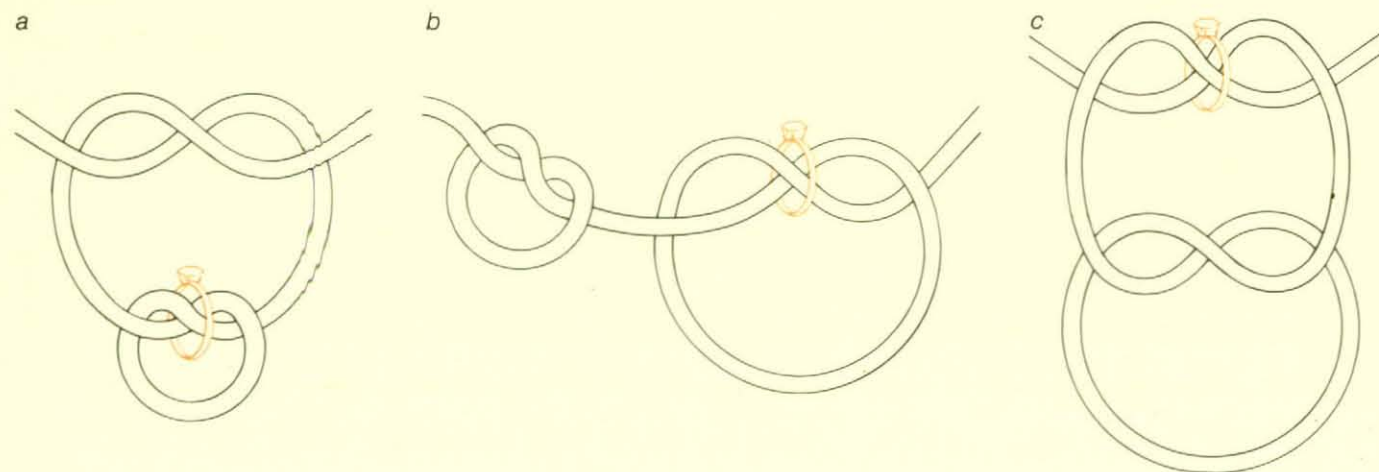
È vero? Una matrice di formule interdipendenti può servire a calcolare tutto ciò che è calcolabile? Il meccanismo è non solo generale ma universale? La domanda ha già avuto una risposta, per il caso di una matrice infinita. Conway ha dimostrato che il mondo cellulare del gioco Vita ha risorse sufficienti per costruire una macchina di Turing, il modello concettuale di un calcolatore universale. Dato che lo si potrebbe usare per giocare a Vita, si potrebbe impiegare un tabellone infinito anche per creare la macchina di Turing.

Tale risultato è sicuramente degno di nota, ma, anche se si può trascurare il requisito di avere un'area infinita, la dimostrazione non ha nessun significato pratico: la vita è troppo breve e il gioco Vita troppo lungo. Io trovo più promettente una impostazione meno formale per stabilire la portata dei programmi di tabellone. È il metodo («o la va o la spacca») di applicare i programmi a vari problemi ed esercizi scelti nel repertorio classico della scienza del calcolatore. I casi interessanti sono verosimilmente quelli con una soluzione algoritmica molto efficiente. Ne è un esempio la Torre di Hanoi, in cui parecchi dischi sono impilati in ordine decrescente su uno dei tre perni; scopo del gioco è spostare i dischi uno per volta, evitando di mettere un disco più grande sopra uno più piccolo, finché non risultino impilati nello stesso ordine su un altro perno. La soluzione standard si basa su un algoritmo ricorsivo che stabilisce esplicitamente lo stadio finale della procedura e poi si basa su di esso per definire gli stadi precedenti.

Si può risolvere la Torre di Hanoi con metodi completamente non algoritmici? Si può fare con un tabellone? Arrivare a tale soluzione non costituirebbe sicuramente una dimostrazione che un tabellone possa fare tutto quello che può fare un algoritmo, ma ne estenderebbe considerevolmente il campo d'azione. Si noti che esiste un metodo di soluzione banale che non è ammesso. Si può risolvere manualmente il gioco, annotando la configurazione dei dischi a ogni stadio e poi scrivere una serie di formule che specifichino la transizione da una configurazione all'altra. È caratteristico di tali metodi forzati che, non appena si opera un piccolo cambiamento nelle condizioni iniziali, quale può essere rappresentato dall'aggiunta di un altro disco, si deve ricominciare. Una buona soluzione dovrebbe rimanere valida per raggruppamenti di qualsiasi entità senza cambiamenti o perlomeno soltanto con cambiamenti nelle dimensioni della matrice. Si noti anche che almeno un programma di tabellone, 1-2-3 della Lotus, comprende un semplice linguaggio algoritmico. Ovviamente, anche questo strumento deve essere bandito.

Un altro caso interessante è il problema delle otto regine, in cui si tratta di disporre otto regine su una scacchiera standard in modo che nessuna regina sia sotto attacco. In questo caso il «formato» del problema - la matrice finita delle caselle - è allettante. Non c'è nessuna difficoltà a rappresentare una scacchiera con un programma di tabellone, ed è facile anche scrivere una formula che indichi se una casella è sotto attacco di una regina posta in qualsiasi punto della scacchiera. (La formula si limita a verificare se c'è un valore diverso da zero lungo tutte le righe, le colonne e le diagonali per una distanza di otto caselle.) Se fosse tutto qui, tuttavia, il problema non avrebbe attirato l'attenzione di Carl Friedrich Gauss, che lo esaminò nel 1850, ma non riuscì a risolverlo. Sembra che per ogni casella sia necessaria l'informazione non solo sull'attuale configurazione della scacchiera, ma anche sulle configurazioni precedentemente sperimentate. La difficoltà di fornire un'informazione come questa in una rappresentazione statica del problema fa pensare che gli algoritmi abbiano un futuro sicuro.

Ecco le soluzioni ai due problemi sui nodi posti il mese scorso da Martin Gardner. La figura in alto in questa pagina mostra come si può trasformare un nodo piano in un nodo alterno a sei incroci. Basta ripiegare l'arco tratteggiato *a* in modo da ottenere l'arco *b*.

Nella figura in basso si può vedere un modo per risolvere il rompicapo dell'anello e del nodo d'asino. Dapprima si rimpicciolisce il nodo in basso e poi lo si fa scivolare insieme all'anello in alto facendolo passare attraverso il nodo più in alto (*a*). Lo si apra e si avranno due nodi a trifoglio uno di fianco all'altro (*b*). Si rimpicciolisca il nodo senza l'anello e lo si faccia scivolare verso il basso passando attraverso l'altro nodo. Lo si apra e il gioco è finito (*c*).

Soluzione del rompicapo dell'anello e del nodo d'asino